



جامعة اليرموك
كلية التربية
قسم علم النفس الإرشادي والتربوي

دقة تقدير معالم الفقرات بطريقتي الأرجحية العظمى الهامشية وبييز في ظروف مختلفة في عدد الفقرات وحجم العينة والنموذج اللوغاريتمي المستخدم

Accuracy of Estimating item Parameters Via Marginal Maximum Likelihood and Bayesian Methods in accordance with difference in Number of Items, Sample size and Logistic Models

إعداد
شادي يوسف خلف الشواورة

إشراف
الأستاذ الدكتور ساري سليم سواقد

حقل التخصص _ قياس وتقويم
الفصل الأول: 18 /2013/12م

دقة تقدير معالم الفقرات بطريقتي الأرجحية العظمى الهامشية وبييز في ظروف مختلفة في عدد الفقرات وحجم العينة والنموذج اللوغاريتمي المستخدم.

إعداد

شادي يوسف الشواورة

قدمت هذه الاطروحة استكمالا لمتطلبات الحصول على درجة دكتوراه الفلسفة في تخصص القياس والتقويم في جامعة، اليرموك، اربد، الاردن.

أعضاء لجنة المناقشة

الاستاذ الدكتور ساري سليم سوالمة.. مشرفاً ورئيساً

استاذ القياس والتقويم، جامعة مؤتة

الاستاذ الدكتور أحمد سليمان عودة .... عضواً

أستاذ القياس والتقويم، جامعة اليرموك

الاستاذ الدكتور يوسف محمد السوالمه .... عضواً

أستاذ القياس والإحصاء التربوي، جامعة اليرموك

الاستاذ الدكتور أحمد يوسف قواسمة .... عضواً

أستاذ القياس والتقويم، جامعة اليرموك

الدكتور محمود فيصل القرعان .... عضواً

استاذ مشارك في القياس والتقويم، جامعة الحسين بن طلال

تاريخ مناقشة الأطروحة : 2013/12/18

الاهداء

إلى من علمني القيم والمبادئ

والدي. . . . . رحمة الله عليه

إلى والدتي الغالية. . . . . أطال الله في عمرها

إلى رفيقة الدرب. . . . . زوجتي العزيزة

إلى قرة عيني. . . راشد ومحمد ورند وفاطمة وميس

إلى إخواني وأخواتي. . . . . . . وفقهم الله

الباحث

شادي يوسف الشواورة

شكر وتقدير

الحمد لله الذي منّ عليّ من فضله وتوفيقه ونعمه لإنجاز هذا العمل المتواضع، وأصلي وأسلم على نبيه سيدنا محمد وصحبه أجمعين، وأسال الله، أن يكون هذا العمل في ميزان حسناتي، وأن يكون مصدر نفع وفائدة لغيري. وأجد لزاما عليّ في هذا المقام أن أنسب الفضل إلى أهله، وأن اتقدم من الأستاذ الدكتور ساري سليم سواقد بجزيل الشكر والعرفان، لإشرافه على هذه الاطروحة، فأغناها بعلمه ورعايته ونصحه، وغمرني بخلقه الكريم فكان نعم المعين، فأسأل الله أن يحفظه ويمتعه بالصحة والعافية. ولا يفوتني أن أشكر اساتذه القياس والتقويم الذين نهلنا من علمهم وحكمتهم اثناء تواجدنا على مقاعد الدراسة في جامعة اليرموك، كما يطيب لي أن اتقدم بجزيل الشكر وعظيم الامتنان والتقدير لأعضاء لجنة المناقشة الأستاذ الدكتور أحمد سليمان عودة، والاستاذ الدكتور يوسف محمد السوالمه، والاستاذ الدكتور أحمد يوسف قواسمة، والدكتور محمود فيصل القرعان، لتفضلهم بقبول مناقشة هذه الاطروحة وإبداء الملاحظات والتوجيهات حولها. وأخيرا أسأل الله التوفيق والصلاح في محياي ومماتي.

فهرس المحتويات

فهرس الجداول

فهرس الأشكال

فهرس الملاحق

الملخص

الشواورة، شادي يوسف، دقة تقدير معالم الفقرات بطريقتي الأرجحية العظمى الهامشية وبييز في ظروف مختلفة في عدد الفقرات وحجم العينة والنموذج اللوغاريتمي المستخدم. أطروحة دكتوراه، جامعة اليرموك 2013. إشراف: أ. د ساري سواقد.

هدفت هذه الدراسة إلى بيان أي من طريقتي التقدير (الأرجحية العظمى وبييز) أكثر دقة في تقدير معالم الفقرات تحت أي من المواقف المتعلقة باختلاف حجم العينة وعدد الفقرات والنموذج اللوغاريتمي المستخدم، ولتحقيق هذا الهدف تم توليد بيانات ثنائية الإجابة، وفقاً للنموذج اللوغاريتمي المستخدم (1-PL,2-PL)، بحجم عينة (1500,1000,500,250)، وعدد فقرات (40,10)، بعد ذلك تم تقدير معالم الفقرات تحت كل موقف من المواقف السابقة (حجم العينة وعدد الفقرات والنموذج) باستخدام طرق تقدير معالم الفقرات: الأرجحية العظمى في التقدير باستخدام برنامج (Bilog-MG 3)، وطريقة بييز في التقدير باستخدام برنامج (WinBUGS V1.4). وتم تحليل قيم معالم الفقرات المقدرة وفقاً لطريقتي التقدير (الأرجحية العظمى وبييز) بواسطة مؤشرات دقة التقدير ومنها: مؤشر الدقة (الجذر التربيعي لمعدل مربعات الأخطاء (RMSE)، ومعامل الارتباط بين قيمتي المعلمة: الحقيقية والمقدرة وفقا لطريقتي التقدير (الأرجحية العظمى وبييز) وبعد ذلك أجري اختبار (t-test) للمقارنة بين معاملي الارتباط لطريقتي التقدير، ومؤشر الكفاءة النسبية (Relative Efficiency). وبعد أن تمت عملية التحليل لوحظ من نتائج مؤشر الدقة (RMSE) أفضلية طريقة بييز في التقدير. وهذا ما أكدته نتائج اختبار (t-test) للمقارنة بين معاملي الارتباط لطريقتي التقدير، حيث أظهرت معاملات الارتباط لطريقة بييز فروقا دالة إحصائياً عند مستوى دلالة (α=0.05) في معظم مواقف الدراسة.

كما بينت نتائج الدراسة أنّ طريقة Bayes للتقدير أظهرت دقة في تقدير معالم الفقرات مقارنة مع طريقة الأرجحية العظمى، وخاصة في العينات الصغيرة، أو عند استخدام النموذج اللوغاريتمي الثنائي (2-PL).

الكلمات المفتاحية: الأرجحية العظمى، طريقة بييز، دقة التقدير، حجم العينة، عدد الفقرات، النموذج اللوغاريتمي.

الفصل الأول

المقدمة

يعرَّف القياس النفسي والتربوي بأنَّه تعيين أرقام أو رموز، تناظر خصائص أو سمات الأفراد طبقًا لقواعد محددة تحديداً جيداً، وهذا يعني أن القياس التربوي والنفسي يُعنى بتصنيف السمات النوعية وتكميم السمات الكمية للأفراد (علام، 2006). ويتصف القياس النفسي والتربوي بمجموعة من الخصائص منها أنه قياس غير مباشر، حيث يتم الاستدلال على السمة المقيسة من خلال ملاحظة السلوكات الدالة وقياسها، فيما نجد أن القياس الفيزيائي الذي يمكن إجراؤه بطرق مباشرة، مما يجعله أكثر دقة وموضوعيةً من القياس النفسي والتربوي الذي يرافقه احتمال الوقوع بالخطأ.

لأجل ذلك ظهرت نظريات القياس النفسي والتربوي، للوصول إلى مؤشرات عن الدقة في القياس، *تشتمل كل نظرية على* مجموعة من الافتراضات، يتم في ضوئها تحديد طرق لتقدير دقة القياس النفسي والتربوي، التي تساعد في اتخاذ القرارات الصائبة في المجالات المختلفة في ضوء نتائج أدوات القياس التي بنيت من اجل قياس السمة المعنية بدرجة عالية من الدقة.

ويوجد في الوقت الحاضر أطر تستند إلى أسس احصائية للتعامل مع مشكلات القياس أهمها: النظرية الكلاسيكية (Classical Test Theory, CCT) ونظرية استجابة الفقرة (Item Response Test theory, IRT)، ويتم الإشارة إلى النظرية الكلاسيكية في القياس (CCT) على أنها "نماذج ضعيفة" ويعزى ذلك إلى أن افتراضات النظرية يمكن تحقيقها بسهولة بواسطة بيانات الاختبار بينما يشار إلى نماذج نظرية استجابة الفقرة على أنها "نماذج قوية" مستمدة قوتها من افتراضاتها التي لا يمكن تحقيقها بسهولة بواسطة بيانات الاختبار.

إن انتشار النظرية الكلاسيكية وسيطرتها على تحليل نتائج أدوات القياس لفترة طويلة من الزمن لم يخفِ بعض جوانب القصور فيها، فهي تفترض تساوي تباين أخطاء القياس لجميع الأفراد الذين يطبق عليهم الاختبار أو المقياس. (Hambleton & Swaminathan ,1985)

جاءت التطورات الحديثة في نظرية القياس، للتأكيد على الموضوعية في القياس وضرورة تقدير الخطأ المعياري للقياس لكل مستوى من مستويات السمة المقيسة. ففي نظرية استجابة الفقرة (IRT) يتم تحديد خطأ القياس عند كل قيمة من قيم القدرة (θ) وعند كل قيمة من قيم معالم الفقرة. نظرية استجابة الفقرة تقوم على افتراضات انبثق منها معادلات (نماذج) لتقدير معالم الفقرة والأفراد يمكن تمثيلها بما يسمى (منحنى خصائص الفقرة) وتحتاج هذه التقديرات إلى معالجات رياضية معقدة ولذلك ظهرت طرق لحل هذه المعادلات وبالتالي تقدير معالم الفقرات والأفراد في النماذج المختلفة لنظرية استجابة الفقرة منها: الأرجحية العظمى للتقدير (Maximum Likelihood Estimation)، ثم ظهرت طريقة بييز للتقدير Bayes) (Estimation، ويرافق كل منهما قيمة للخطأ المعياري لتقدير قيم القدرة (θ) ومعالم الفقرة (الصعوبة، والتمييز، والتخمين). وبما أنه لا يوجد إجماع على أفضلية أي من الطريقتين أدق في تقدير معالم في نظرية استجابة الفقرة، جاءت هذه الدراسة، لمعرفة أي الطريقتين (الأرجحية العظمى، وطريقة بييز) تعطي أفضل دقة في تقدير معالم الفقرة، وفي ضوء بعض المتغيرات مثل حجم العينة وطول الاختبار والنموذج اللوغاريتمي (اللوجستي) المستخدم، واهتمت هذه الدراسة بمعالم الفقرات وذلك لأهميتها في تقدير قدرة الأفراد، ولكي تكون تقديرات القدرة على درجة عالية من الدقة فلا بد من مراعاة دقة تقدير معالم الفقرة.

الأدب النظري:

إن القياس النفسي والتربوي يهتم بدراسة السمات وقياسها لدراسة السمات وقياسها، ويعد تحديد السمة المقيسة وتعريفها من الخطوات الأساسية في قياسها، والسمة مفهوم مجرد يُحدد أنماطًا معينةً من السلوك، وبالتالي فهي عبارة عن تجمع من السلوكات المترابطة التي يحتمل حدوثها معاً، فالسمة ليست شيئاً ملموساً، والسمات لا يمكن قياسها بشكل مباشر، وإنما يستدل عليها من أنماط السلوك الملاحظ. ويجب أن تفي السمات ببعض الشروط لكي يتسنى لنا الإفادة منها في وصف السلوك منها أن تمثل السمة خاصة معينة يتباين فيها الأفراد، وأن تكون السمة قابلة للتعريف الإجرائي، وأن تكون السمة متسقة بدرجة معقولة عبر الزمن أي أن السمة المعنية يجب ألا تتغير لدى الفرد من وقت لآخر ، ولكي تكون السمة خاصية محددة يتباين فيها الأفراد، وأن تكون السمة قابلةً للتعريف الإجرائي وتتصف بالثبات النسبي (علام، 2006).

بدأت دراسة هذه السمات وقياسها، من خلال نظرية الاختبار في القياس التي أظهرت بعض القصور في إجراءاتها. حيث تفترض نظرية الاختبار في القياس عدم وجود ارتباط بين أخطاء القياس لعدد من المفحوصين في الاختبار نفسه، وعدم وجود ارتباط بين أخطاء القياس في أي اختبارين مطبقين على المفحوص، وإن خطأ القياس هو خطأ عشوائي، وكذلك تجانس تباين الخطأ في القياس. ومن جانب أخر تستدل هذه النظرية على دقة القياس من خلال معامل الثبات (Crocker & Algina, 1986)

يعرّف ثبات الاختبار بأنه قدرة الاختبار على قياس السمة بدقة (أي مقدار الدقة في القياس) ويعبر عن مقدار الدقة بقيمة رقمية تتراوح نظريًا بين صفر و واحد صحيح، وتشير متممة هذه القيمة إلى مقدار عدم الدقة في القياس، الذي يعبر عنه بمقدار الخطأ المعياري للقياس. يعرّف عودة (2010)، ثبات الاختبار بأنه الاختبار الذي يقيس بدرجة مقبولة من الدقة أو أقل خطأ ممكن، ولابد من الإشارة هنا إلى الخطأ المعياري للقياس (Standard Error of

(Measurement, (SEM))، فإذا كان مقدار هذا الخطأ صفراً، تكون درجات الاختبار ذات ثبات تام، لكن لا يوجد قياس نفسي وتربوي يخلو من أخطاء في القياس – هذا ما تفترضه نظرية الاختبار – مما يؤدي إلى خفض ثبات درجات الاختبار بقيم متفاوتة بحسب مقدار هذا الخطأ. علماً أن قيمة الخطأ المعياري للقياس هو الانحراف المعياري لعلامات الفرد إذا ما طبق الاختبار عليه مرات عدة، وهذا أمر غير ممكن، لذلك فإننا نقدر هذه القيمة إذا علمنا قيمة الانحراف المعياري للدرجات الملاحظة، وكذلك قيمة معامل ثبات درجات الاختبار من خلال استخدام الصيغة الرياضية الآتية: $SEM = \sigma \sqrt{1-r}$ ، حيث أن SEM يرمز للخطأ المعياري في القياس، أما الرمز σ فيعني الانحراف المعياري و r ترمز إلى معامل الثبات.

نماذج نظرية استجابة الفقرة:

انبثقت نظرية استجابة الفقرة من افتراض إمكانية التنبؤ بأداء الأفراد أو تفسير أدائهم في اختبار ما في ضوء خصائص مميزة لهذا الأداء تسمى القدرات، التي لا يمكن ملاحظتها بشكل مباشر وإنما من خلال أداء الأفراد في مجموعة من الفقرات لذلك تهدف نظرية استجابة الفقرة إلى تحديد العلاقة الرياضية الاحتمالية بين أداء الفرد في اختبار ما، والسمة الكامنة وراء هذا الأداء، ويُعبر عن هذه العلاقة في النماذج اللوغاريتمية (اللوجستية)، وهذه النماذج تحاول تقدير موقع الفرد (المفحوص) على متصل السمة أو القدرة باستخدام نمط استجاباته، وتقوم هذه النظرية على افتراض أساسي مفاده أن احتمال استجابة الفرد على أي فقرة من فقرات الاختبار بشكل صحيح، يمثل اقتراناً لكل من قدرة الفرد التي يقيسها الاختبار من جهة وخصائص الفقرة التي يحاول ذلك الفرد الاجابة عنها من جهة ثانية، وهذا يتطلب الحصول على معلومات حول الفرد والفقرة، أي الحصول على تقديرات لقدرات الأفراد ومعالم الفقرات. (Crocker 1986 & Algina,).

أشار بيكر (Baker, 2001) إلى النماذج الرياضية لنظرية استجابة الفقرة التي تختلف باختلاف معالم الفقرات المراد تقديرها، فالنموذج اللوغاريتمي أحادي المعلم الذي اقترحه جورج راش يقدر معلمة الصعوبة وقدرة الفرد، أما النموذج اللوغاريتمي ثنائي المعلم الـذي اقترحـه بيرنبوم يقدر معالم الفقرة (الصعوبة والتمييز) وقدرة الفرد، في حين أن النموذج اللوغـاريتمي ثلاثي المعلم (نموذج لورد) يقدر معالم الفقرات (الصعوبة والتمييز والتخمين) وقدرة الفرد، وهذه النماذج تعتبر من أكثر نماذج نظرية استجابة الفقرة شيوعاً، وأكثرها ملاءمة مع الفقرات ثنائية التدريج، وتاليا توضيح للنماذج اللوغارتمية الثلاث:

النموذج اللوغاريتمي أحادي المعلم (One-Parameter Logistic Model):

سعى واضع هذا النموذج العالم الدنماركي جورج راش في عام 1960 إلـى تحقيـق الموضوعية في القياس النفسي التربوي من خلال توصله إلى أدوات قياس اتسمت باسـتقلاليتها عن خصائص الأفراد المراد قياس قدرتهم، فضلاً عن استقلالية قدرة الأفراد المقاسة عن أدوات القياس، إلا أن هذا النموذج لم يرَ النور إلا بعد تطبيقه مـن العـالم الأمريكـي (Berjamin Wright) عام 1967 حيث عرض دراسةً توضح كيف تنتقل من الجانب الرياضـي النظـري للنموذج الذي قدمه راش إلى الجانب التطبيقي (Hambleton & Cook, 1977). ويهتم هذا النموذج بتحديد موقع الفقرة الاختبارية على تدريج صعوبة جميع الفقرات التي تشكل الاختبار، كما يهتم بتدريج مستويات قدرة الفرد باختبار معين على نفس التدريج.

ومن الجدير ذكره أن هذا النموذج يفترض تساوي قوة تمييز جميع فقرات الاختبـار وأن التخمين لجميع الفقرات في حدودها الدنيا أو ما دونها، ويعبر عن الدالة الرياضـية لهـذا النموذج كما يلي:

$$p_i(\theta)=\frac{e^{D\bar{a}_i(\theta-\mathrm{bi})}}{1+e^{D\bar{a}_i(\theta-\mathrm{bi})}} \qquad ....(1)$$

Pi(θ) هي احتمال الإجابة الصحيحة للفرد j الذي قدرته jΘ على الفقرة i التي صـــعوبتها bi حيث إن (i=1 , 2,.........., n) و n هي عدد الفقرات. وأن D هــو معامــل التــدريج اللوغاريتمي الذي يساوي واحد صحيح. e هي الدالة الأسية ذات الأساس الطبيعي (2.71828) و D ثابــت التــدريج (Sceling Fector) ويســاوي تقريبــاً (1.7). (Koning, E. D., Sijtsma, K. & Hamers, J. H. M. 2002).

النموذج اللوغاريتمي ثنائي المعلم (Two-Parameter Logistic Model):

النموذج ثنائي المعلم يسمح باختلاف فقرات الاختبار من حيث صــعوبتها وتمييزهــا ويمثل بالدالة الرياضية الآتية:

$$p_i(\theta)=\frac{e^{Dai(\theta-\mathrm{bi})}}{1+e^{Dai(\theta-\mathrm{bi})}} \qquad ....(2)$$

إذ أن pi(θ) هو احتمال اجابه المفحوص الذي قدرته jθ على الفقرة i التي صعوبتها bi وتمييزها ai حيث (i=1 2,......,n) و n عــدد الفقــرات فــي الاختبــار (Borsboon, D . (Ferrando, P. J. , Lorenzo - ،Mellenbergh, G, & Heerden, J.V. 2002). Seva, U. , & Molina, G. 2001). ويلاحظ في هذه المعادلة الرياضية أن معامل التمييــز للفقرة يضرب بالفرق بين مستوى صعوبة الفقرة ومستوى القدرة ويكون تأثير هذا الفــرق علــى احتمال الإجابة الصحيحة مرتبطًا بالقوة التمييزية للفقرة، حيث يكون تأثير معلمة التمييــز للفقــرة كبيرًا عندما تكون القيمة التمييزية للفقرة ذات قيمة كبيرة، ونلاحظ أن القيمة التمييزية للفقرة فــي

النموذج ثنائي المعلمة لا يكون لها نفس الوزن النسبي في تقدير مستويات القدرة (Embretson, Reise 2000).

النموذج اللوغاريتمي ثلاثي المعلم (Three-Parameter Logistic Model):

من المشكلات التي واجهت النموذج اللوغاريتمي ثنائي المعلمة ان هنالك احتمالاً لإجابة الأفراد ذوي القدرة المنخفضة على الفقرة، وتكون نسبة إجابتهم أكبر من الصفر وذلك بسبب عامل التخمين في الإجابات عن الفقرات الاختبارية. لذلك قام النموذج اللوغاريتمي الثلاثي على ثلاثة معالم وهي معلمة الصعوبة (bi) ومعلمة التمييز (ai) ومعلمة التخمين (Ci)، ويحدد هذا المعلم على منحنى خصائص الفقرة في الجزء الأسفل منه ويدعى خط التقارب الأدنى (Lower Asymptote) (Glas, Falcon 2003, Syker, Hou ,2003). ويصاغ النموذج الثلاثي على النحو التالي

$$p_i(\theta)=Ci+(1-Ci)\frac{e^{Dai(\theta-bi)}}{1+e^{Dai(\theta-bi)}} \quad .... (3)$$

حيث إن:

$pi(\theta)$: احتمال إجابة المفحوص الذي اختير عشوائياً من مستوى القدرة (θ)عن الفقـرة (i) إجابةً صحيحة، bi: معلمة الصعوبة، ai: معلمة التمييز، Ci: معلمة التخمين، θ: معلمة القدرة D: عامل التدريج (Scaling Factor).

وعند النظر في النماذج الثلاثة في نظرية استجابة الفقـرة (النمـوذج اللوغـاريتمي الأحادي، والنموذج اللوغاريتمي الثنائي، والنموذج اللوغاريتمي الثلاثي) نجد أن النموذج أحادي المعلمة هو حالة خاصة من النموذج ثنائي المعلم، عندما تكون قيم كل من معلمة التمييز واحـد ($a_i=1$) والنموذج ثنائي المعلم حالة خاصة من النموذج ثلاثي المعلـم عنـدما تكـون معلمـة التخميين صفر (Ci=0) لجميع الفقرات.

دقة تقدير معالم الفقرة:

عملية تقدير معالم الفقرة تتطلب دقة في التقدير، وبأقل خطأ تقدير ممكن، ويتوقف الخطأ على الإجراءات المتبعة في تقدير معالم الفقرة، فقد أشار هامبلتون وسامبينثان (Hambelton & Sweminthen, 1985)، وعودة والخليلي (2000) وعلام (2005) ولورد (Lord,1986) إلى محكات ينبغي الاستناد عليها في الحكم على دقة التقدير وتقييم مدى جودة الأسلوب الإحصائي المستخدم للوصول إلى قيم تقديرية للمعلم (البارميتر) وهي:

1-عدم التحيز (Unbiasedness):

يعني أن القيم المقدرة لمعالم الفقرة وفق الطريقة المستخدمة قريبة من القيم الحقيقية لتلك المعالم، وللوصول لهذا المؤشر يجب إيجاد الفرق بين القيم المقدرة والحقيقية.

2- معدل مربعات الأخطاء للتقدير (Mean Squared Error : MSE):

يتم قياس هذا المؤشر من خلال تباين القيمة المقدرة لمعالم الفقرات، وبالاعتماد على مؤشر التحيز يمكن معرفة الفرق بين القيمة المقدرة والقيمة الحقيقية، باستخدام مؤشر الجذر التربيعي لمعدل مربع الخطأ (Root Mean Square Error:RMSE)، وكذلك الفرق بين مقدرين.

3- الاتساق (consistency):

يقصد بالاتساق أن التوقع في قيمة الإحصائي المستخرج بالمعاينة يقترب أكثر من قيمة المعلم في المجتمع الإحصائي، بازدياد حجم العينة، وهذا ما أشار إليه كل من هامبلون وكوك (Hambleton & Cook ,1983) ولورد (lord ,1980) حول أثر زيادة عدد الأفراد والفقرات في اتساق تقديرات معالم الفقرات الثنائية ودقتها.

4-الكفاءة النسبية (Relative Efficiency):

تشير الكفاءة إلى الدقة التي يقدّر بها إحصائي معين المعلم المناظر لـه فـي المجتمـع الإحصائي، وهي تشير إلى التشتت في التقدير من عينة إلى أخرى، وتعني إذا توفر مقياسـان إحصائيان غير متحيزين لتقدير معلم المجتمع فإن أفضلهما هو الاحصائي الذي يكـون أكثـر فاعلية بالنسبة للآخر، وتحتكم إلى الخطأ المعياري لتوزيع المعاينة الأقل.

طرق تقدير معالم الفقرات:

إن تقدير معالم الفقرات يعتبر ركيزةً أساسيةً نستطيع من خلالها تقدير قدرة الأفـراد وفق نظرية استجابة الفقرة، ونظراً إلى أن معالم الفقرات غير معروفةً فإننـا نقـوم بتقـديرها باستخدام أحد طرق التقدير، وذلك بالاعتماد على إجابات الأفراد على فقرات الاختبار، وتحتاج عملية تحديد معالم الفقرات في نماذج نظرية استجابة الفقرة إلى دقة في التقدير وبأقل خطأ تقدير ممكن، ويتوقف ذلك على الإجراءات المستخدمة في عملية التقدير لمعالم الفقرة.

يوجد عدة طرق لتقدير معالم الفقرة وفق نظرية استجابة الفقرة فقـد أشـار هـامبلتون وسوامينثان (Hambeleton & Swaminithan, 1985) إلى الطرق الآتية:

طريقة الأرجحية العظمى للتقدير (Maximum Likelihood Estimation):

وهي من الطرق الرئيسية في تقدير معالم الفقرة والقدرة ويتم ذلك من خلال إجـراءات تعظيم الاحتمالية للمعلمة المراد تقديرها، وهذه الطريقة مستخدمة في برنامج (Bilog. Mg 3) فقد بين فرانك ريجيمين (Frank, Rijmen, 2009) أن هذه الطريقة تقوم على إيجاد تقـدير المعالم من خلال إجراءات تعظيم الاحتمالية للمعلمة المراد تقديرها عندما يتوافر لدينا معلومات عن العينة. ومن أساليب الأرجحية العظمى في تقدير معالم الفقرات ما يأتي:

1) الأرجحية العظمى المشتركة (Joint Maximum Likelihood: JML):

إحدى طرق الأرجحية العظمى في تقدير معالم قدرة الفقرات وتتم عمليه التقدير وفق هـذه الطريقة في مرحلتين:

- **في المرحلة الأولى:** يتم اختيار قيمة مبدئية لمعلمة الأفراد (θ) ويتم احتساب هذه القيمة المبدئية من خلال النسبة بين الإجابات الصحيحة والإجابات الخاطئة لكل فرد (مختبر) والناتج يتم تحويله إلى قيم معيارية وذلـك للـتخلص مـن مشـكلة عـدم التحديـد (Indeterminacy).
- **المرحلة الثانية:** يتم التعامل مع معالم الفقرات كما تعاملنا مع معلمة القدرة (أي أن معالم الفقرات معلومة).

يتم تتابع (تكرار) هاتين المرحلتين حتى نصل للقيم التقديرية للمعـالم التـي لا تتغيـر بتكرار عملية التقدير (التقي، 2010)، (عـلام، 2005)،(Hambleton & Swaminathan, 1985). ووفق طريقه (JML) هنالك محكات تُعتمد للتوقف عن عملية التتابع (التكرار) لتقدير معالم الفقرة وهي:

المحك الأول: الوصول إلى قيم تقديرية لمعالم الفقرات لا تتغير بعد الانتهـاء مـن مـرحلتين متتاليتين أو أن التغير قليلا يمكن إهماله، حيث نكون قد وصلنا إلى الثبـات فـي عمليـة التقدير.

المحك الثاني: أن يكون الفرق في التقديرات في أي من المراحل أقل من عدد (محـك) محـدد مسبق.

إن تقدير معالم الفقرات يعتمد على البدء بقدرة (θ) معينه لكل الأفراد التي يقدر مـن خلالها معالم الفقرات (0β) التي يعتمد عليها لاحقا في تقدير قدرة الأفراد، وتتكرر هذه العمليـة

حتى نصل إلى المحك المعتمد للتوقف عن عملية التكرار. والمعادلة التي قدمها هامبلتون وسوامينثان(Hambleton &Swaminathan ,1985) توضح كيفية استخراج قيم معالم الفقرات (a, b, c) ومعلمة الأفراد (القدرة، θ) في النموذج اللوغاريتمي ثلاثي المعلم وفق أسلوب (JML)

$$In\,\mathrm{L}\,(\mathrm{u}|\theta,b,a,c)=\sum_{a=1}^{N}\sum_{i=1}^{n}\left[u_{ia}InP_{ia}+(1-u_{ia})InQ_{ia}\right] \qquad ....(4)$$

حيث أن (u) متجه ذو (Nn) من الأبعاد، لاستجابة (N) من الأفراد على (n) من الفقرات، أما متجهات (Θ, b, a, c) فهي متجهات تضم معالم (القدرة والصعوبة والتمييز والتخمين)، تشير p_{ia} إلى الإجابات الصحيحة اما Q_{ia} فإنها ترمز إلى الإجابات الخاطئة.

وعند تطبيق هذه المعادلة على النموذج اللوغاريتمي الأحادي (راش) أي استخراج معلمة الصعوبة تصبح المعادلة السابقة على الصورة الآتية:

$$In\,\mathrm{L}\,(\beta|\theta s,uis)=\sum_{s=1}^{N}\left[u_{ia}InP_{ia}+(1-u_{ia})InQ_{ia}\right] \qquad ....(5)$$

فإذا أجاب (5) مفحوصين، قيمة السمة الكامنة لهم هي: (1,1,0,0,-1) على التوالي، ومثّل المتجه (0,1,1,1,1) الإجابة على هذه الفقرة وتكون بالصورة الآتية:

u=(u1,u2,u3,u4,u5) = (0,1,1,1,1)

$$L(u_1,u_2,...,u_n|\theta)=\prod_{i=1}^{n} P_i^{u_i}Q_i^{1-u_i} \qquad ....(6)$$

$$L(u|\theta)=Q_1P_2Q_3P_4P_5$$

$$P_i=\exp\mathrm{D}(\theta-\mathrm{b_i})/[1+\exp\mathrm{D}(\theta-\mathrm{b_i})] \qquad ....(7)$$

$$Q_i = 1/[1+\exp D(\theta - b_i)] \qquad ....(8)$$

وفق أسلوب نيوتن رافسون الذي يقوم على حساب المشتقة الأولى والمشتقة الثانية لاقتران لوغريتم للأرجحية العظمى، ويتم احتساب المشتقة الأولى بالنسبة لمعلمة الصعوبة في النموذج الأحادي التي تكون عندها المشتقة الأولى تساوي صفرًا.

للوصول إلى المشتقة الأولى يكون وفق المعادلة الآتية:

$$l'(\beta_i) = \frac{\partial}{\partial \beta_i} l(\beta_i | \ (u_i, \theta_s) = \sum_{s=1}^{N} [u_{is} - P_{is}] \qquad ....(9)$$

المشتقة الثانية تحسب وفق المعادلة الآتية:

$$l''(\beta_i) = \frac{\partial^2}{\partial \beta_i^{\,2}} l(\beta_i | \ (u_i, \theta_s) = \sum_{s=1}^{N} Pis[1 - P_{is}] \qquad ....(10)$$

ويتم إيجاد القيمة العظمى لأي اقتران بإيجاد قيمة المتغير (المعلم) الذي يجعل المشتقة الأولى عنده تساوي الصفر، وبالتالي فإن إيجاد درجة صعوبة الفقرة (من الفقرة 1 إلى الفقرة L) من خلال تطبيق المعادلة الآتية:

$$\sum_{s=1}^{N} u_{is} - \sum_{s=1}^{N} P(u_{is} = 1 | \ r_s \beta_j) = 0 \qquad ......(11)$$

ويستخدم أسلوب نيوتن رافسون في ذلك حيث يتم ما يلي لكل من المعادلات أعلاه:

1. يتم افتراض قيمة (صفر) لمعلمة الصعوبة لتشكل هذه القيمة (الحاليةβ) التي يتم تقديرها في أي من الدورات. $f\ (X) = 0$
2. بالاعتماد على قيمة الصعوبة (صفر) يتم إيجاد المشتقة الأولى $(r\beta_i)$، والمشتقة الثانية $(l''\beta_i)$ لكل من قدرات الأفراد (θ) التي تكون معلومة.

3. يتم إيجاد حاصل قسمة المشتقة الأولى على المشتقة الثانية ($r\beta_i$) وفق المعادلة الآتية:

$$\varepsilon = \frac{f'(\beta i)}{f''(\beta i)} = h = \frac{f'(x_o)}{f''(x_o)} \quad \text{.......} (12)$$

4. يتم تحسين قيمة معلمة الصعوبة للدورة الآتية من خلال المعادلة الآتية:

$$\beta_{\text{الجديدة}} = \beta_{\text{الحالية}} \cdot \varepsilon \quad \text{......}(13)$$

$$x_1 = x_o - \frac{f'(x_o)}{f''(x_o)} \quad \text{.......} (14)$$

$$x_1 = x_o - h$$

5- القيام بعملية التتابع بالتقدير (تكرار) من خلال وضع القيمة ō(الجديدة β) مكـان القيمـة (الحالية β) وتعاد الخطوات من (2) إلى (4) حتى نصل إلى المحك وعندها تكـون القيمة (الجديدة β) هي القيمة المقدرة لمعلمة الصعوبة (Hambleton &Swaminathan ,1985).

ما يميز هذه الطريقة (JML) أنها لا تفترض توزيعات للقدرة أو معالم الفقرات، إلا أنه يوجـد سلبيات لهذه الطريقة وقد أشار لها (علام، 2005) نذكر منها:

- يعاب على طريقة (JML) أن تقديراتها لمعالم الفقرات تكون متحيزة وفقا لمحكات تقييم دقة التقدير سابقة الذكر.
- التقديرات غير ثابتة ويعود ذلك إلى أن طريقة (JML) تقدر معالم الفقـرات والأفـراد معا، وللتغلب على هذه المشكلة ينبغي تقدير معالم هذه الفقـرات دون الاسـتناد إلـى تقديرات الأفراد، لذا جاءت طريقة الأرجحية العظمى الهامشية (MML) لتعالج مشكلة عدم الثبات وفق طريقة (JML).

- عدم وجود قيمة عظمى للوغاريتم الأرجحية في الحالات التي تكون فيها الإجابة على جميع الفقرات صحيحه، أو العكس أي تكون جميع الفقرات إجاباتها خاطئة، مما يترتب على ذلك عدم إيجاد التقديرات لجميع الفقرات والقدرات، بسبب استبعاد الفقرات التي أجاب عنها الأفراد جميعهم بشكل صحيح، أو أجابوا عنها بشكل خاطئ.
- الأخطاء المعيارية التي يتم احتسابها باستخدام (JML) ليست دقيقة.
- ثبات التقدير يحتاج إلى عدد كبير من الأفراد مما يترتب على ذلك زيادة في عدد القدرات التي يراد تقديرها وبالتالي قد لا تحسن بالشكل المقبول عملية التقدير.

2) طريقة الأرجحية العظمى الشرطية (Conditional Maximum Likelihood: CML):

يعود الفضل في تقديم هذه الطريقة لتقدير معالم الفقرات لاندرسون عام 1972م، ويقتصر استخدام هذه الطريقة على نموذج راش الأحادي (1-PL)، وهذا الأسلوب في تقدير معالم الفقرات يمكن إجراؤه لتقدير (40) فقرة بشكل فعّال، وعند استخدام هذا الطريقة لفقرات يزيد عددها عن ذلك تصبح غير مناسبة، وخاصة عندما يكون عدد الفقرات من (100-80) فقرة، (Hambeleton & Swaminithan, 1985) لكن التطورات التي أجريت على طريقة (CML) جعلتها تُقدر معالم فقرات تصل إلى (100) فقرة.

ويمكن إجراء تقدير لمعالم الفقرة باستخدام (CLM) بشكل أكثر دقة من خلال العلامات الخام على الفقرات دون الرجوع إلى قيمه القدرة (θ)، ويقوم هذا الأسلوب في التقدير على شرطٍ أساسٍ في تقدير معلمة القدرة والصعوبة عند الأفراد، وهو مجموع الفقرات التي أجاب عليها الفرد إجابةً صحيحةً (ui)، وتحذف الفقرات التي تكون عليها الإجابة خاطئة ولا بد من الإشارة هنا إلى أن (CLM) تقدر معالم الفقرات بغض النظر عن قدرة الأفراد (Hambeleton & Swaminithan,

(1985. لإيجاد الأرجحية العظمى لصعوبة الفقرة في نموذج راش وفق طريقه (CML) نطبق المعادلة الآتية:

$$\sum_{s=1}^{N} u_{is} - \sum_{s=1}^{N} P(u_{is} = 1 | r_s \beta_j) = 0 \qquad ....(15)$$

ويرمز r_s إلى مجموع العلامات التي يحصل عليها الطالب s ويرمز i للفقرة. وتجدر الإشارة هنا إلى أن هذه الطريقة لا تصلح للنموذج اللوغاريتمي الثنائي (2-PL)، والنموذج اللوغاريتمي الثلاثي (3-PL) حيث تختلف قيم معلمة تمييز الفقرات، ولذلك يعتمد أسلوب (CML) على أي الفقرات التي أجاب عليها الفرد إجابةً صحيحة وليس على جميع فقرات الاختبار.

وقد أشار كل من أمبرسون وريز (Embreson & Reise,2000) وهامبلتون وسوامينثان (Hambeleton & Swaminithan, 1985) و (علام، 2005) إلى عدة فوائد وميزات لهذه الطريقة وهي:

- لا تعتمد طريقه (CML) على توزيعات مسبقة لقدرة الأفراد (θ) في تقدير معلمة الصعوبة
- تحقق مفهوم اللاتغير (Invariance) في تقدير معالم الفقرات عند اختلاف الأفراد ويعود سبب ذلك إلى عدم اعتماد تقدير الفقرات على قدرة الأفراد.
- تمتاز طريقة (CML) بخصائص مطلوبة لدقة التقدير مثل الثبات.

ويشوب هذه الطريقة بعض العيوب منها:

- طريقة (CML) لا تقدر معلمة الصعوبة للدرجات الكاملة على الاختبار وكذلك عدم وجود تقديرات للدرجات الصفرية على الفقرات، مما يترتب على ذلك فقدان بعض من المعلومات.
- اقتصار استخدام طريقة (CML) على النموذج اللوغاريتمي أحادي المعلم (راش) والنماذج المنبثقة عنه، ولا يمكن تطبيقها على النموذج اللوغاريتمي الثنائي والثلاثي.
- تعطي طريقة (CML) تقديرات غير دقيقة عندما يزيد عدد الفقرات عن (100) فقرة، بسبب ظهور صعوبات في التحليل العددي للبرمجيات المستخدمة وفق هذه الطريقة.

3) طريقة الأرجحية العظمى الهامشية (Marginal Maximum Likelihood: MML):

طورت هذه الطريقة في التقديرعلى يد كل من بسوك وتيكن عام 1981م، وقد استخدمت هذه الطريقة لتقدير معالم الفقرات في البرنامج الحاسوبي (Bilog-MG 3)، وما يميز أسلوب (MML) أنها جاءت لتعالج بعض مشكلات الطريقتين السابقتين (JML,CML)، حيث تقوم بتقدير معالم الفقرات للنماذج اللوغارتمية الأحادية والثنائية والثلاثية (1-PL, 2-PL, 3-PL) وتتعامل بفاعلية مع عدد الفقرات سواء كانت قليلة العدد أو كثيرة، وهذا ما عجزت عنه طريقة (CML)، وتتعامل طريقة (MML) مع جميع فقرات الاختبار، سواء أجاب عليها المختبرون أم لم يجيبوا، وقد عالج أسلوب الأرجحية العظمى الهامشية مشكلة عدم الثبات في طريقة (JML) التي تنتج عن تقدير معالم الأفراد والفقرات معاً، وأكثر ما يميز هذه الطريقة أنها تعطي أقل خطأ تقدير لمعالم الفقرات مقارنة بطرق الأرجحية الأخرى. (Hambeleton & Swaminithan, 1985)، (Lo, Liu & sheo,2003).

وقد أشار كل من كروكر والجينا (Crocker & Algina, 1986) إلى أنـــه يمكـــن إيجاد اقتران الاحتمالية الهامشية لتقديرات معالم الفقرات من خـــلال تكامـــل اقتـــران الكثافـــة الاحتمالية على معالم القدرة للأفراد، ثم إيجاد تقديرات معـــالم الفقـــرات (الصـــعوبة والتمييـــز والتخمين) التي يمكن من خلالها تقدير قدرة الأفراد لاحقًا.

وأشار هامبلتون وسوامينثان (Hambeleton & Swaminithan, 1985) إلـــى أن (بول وليبرمان) أطلقا تسمية التقديرات الشرطية، عندما يتم تقدير معالم الفقرات والقدرة معـــاً، وبشكل متزامنٍ، لأنه في هذه الحالة تتم معاملة المتقدمين للاختبار على أنهم مجهولين وثابتين. أما التقديرات غير الشرطية تكون عندما يتم اعتبار المختبرين عينةً عشوائيةً، ويتم إجراء تكامل لدالة الأرجحية على توزيع القدرة (θ)، لذلك يطلق على تقديرات معالم الفقرة مقـــدرات غيـــر شرطية، وقد أدت هذه التسميات (أي الشرطية، غير الشرطية) إلى نوع من الإرباك من قبـــل الباحثين، لذا نبّه أندرسون ومارسن لهذا الإرباك واقترح استخدام مصطلح التقديرات الهامشية.

أشـــار هـــامبلتون وســـوامينثان (Hambeleton & Swaminithan, 1985) إلـــى الإجراءات التي يتم من خلالها تقدير معالم الفقرات وفق طريقة (MML)، التي تمر بمرحلتين وهما:

*** مرحلة التوقع (Expectation Stage):**

يتم توقع عدد الأفراد في كل مستوى من مستويات القدرة (Θ)، التي يرمز لها بالرمز $n_k^{(s)}$، وكذلك توقع عدد الأفراد (j) الذين أجابوا إجابةً صحيحةً عن الفقرات (i) للقـــدرة ذاتهـــا، الذي يرمز له بالرمز $r_{jk}^{(s)}$، والتي يتم إيجادها من متجه إجابات الأفراد (y_i) للفقـــرات التـــي عددها L.

* **مرحلة التعظيم (Maximization Stage) :**

في هذه المرحلة يتم إيجاد معالم الفقـرات $(\delta_1, \delta_2, ...\delta_L)$، ويرمـز لهـا بالرمز Δ، علماً أن قدرة الأفراد تتبع التوزيع الطبيعي، بحيث تساوي قدرة أي مـن الأفـراد، واحدة من القيم الآتية: $(q_k, k = 1,....., k)$ باحتمال يساوي $(\pi_k, k = 1,....., k)$ والتـي يرمز لها بالرمز π، ويشترط في قيم π، Δ أن البيانات الملاحظة يتم الترجيح لهـا، عنـد القيم $n_k^{(s)}$، $r_{ik}^{(s)}$، التي تم توقعها في المرحلة الأولى، عند تتابع التقدير لابد من الوصـول إلـى المحك وهو أن نحصل على قيمتين متتاليتين من قيم الاحتمالات π، واي قيمتين متتاليتين من قيم معالم الفقرة Δ المتناظرة لعدد نفترضه مسبقًا.

أشار هامبلتون وسـوامينثان (Hambeleton & Swaminithan, 1985) ولـورد (Lord,1986)، إلى الإجراءات والمعادلات المستخدمة في تقدير معالم الفقرات وفق طريقـة (MML). والمثال التالي يوضح الطريقة لتقدير معالم الفقرة (الصعوبة والتمييز والتخمين).

إذا كانت المصفوفة الآتية تمثل إجابات (N) من الأفراد على (L) من الفقرات

Persons \ Items	1	2	3	...	L
1	y_{11}	y_{12}	y_{13}	...	y_{1L}
2	y_{21}	y_{22}	y_{23}	...	y_{2L}
3	y_{31}	y_{32}	y_{33}	...	y_{3L}
⋮	⋮	⋮	⋮	⋮	⋮
N	y_{N1}	y_{N2}	y_{N3}	...	y_{NL}

إذا اعطينا الرمز γ للمصفوفة السابقة والرمز iΘ يمثل قدرة الفرد i ، الذي يدل عليـه متجه الإجابة yi لهذا الفرد على الفقرات الاختباريه التي عددها L ويعبر عن البيانات الكليـة كما يأتي:

$$[(y_1, \theta_1),(y_2, \theta_2)......,(y_N, \theta_N)]$$

وعند معرفة قيم كل من π وΔ يصبح التوزيع الاحتمالي لأي من متجهات الاستجابات

للفقرة y يعطى بالعلاقة الآتية:

$$f(y \mid \Delta, \pi) = \sum_{k=1}^{k} f(y \mid q_k \Delta) \pi_k \quad ....(16)$$

ويعبر عن الأرجحية للبيانات الملاحظة بالعلاقة الآتية:

$$\prod_{i=1}^{N} f(y_i \mid \Delta, \pi) = \prod_{i=1}^{N} \sum_{k=1}^{k} f(y_i, q_k \mid \Delta, \pi) \quad .... (17)$$

ويتم استخدام المعادلات الآتية في مرحلة التوقع:

$$n_k^{(s)} = \sum_{i=1}^{N} \frac{\pi_k^{(s)} \prod_{i=1}^{L} P(q_k \mid \beta_j^{(s)})^{y_{ij}} \left[1 - P(q_j \mid \beta_j^{(s)})^{y_{ij}}\right]^{1-y_{ij}}}{\sum_{k=1}^{K} \pi_{k'}^{(s)} \prod_{i=1}^{L} P(q_j \mid \beta_j^{(s)})^{y_{ij}} \left[1 - P(q_j \mid \beta_j^{(s)})^{y_{ij}}\right]^{1-y_{ij}}} \quad ....(18)$$

$$r_{jk}^{(s)} = \sum_{i=1}^{N} \frac{y_{ij} \pi_k^{(s)} \prod^{L} P(q_k \mid \beta_j^{(s)})^{y_{ij}} \left[1 - P(q_j \mid \beta_j^{(s)})^{y_{ij}}\right]^{1-y_{ij}}}{\sum_{k=1}^{K} \pi_{k'}^{(s)} \prod_{j=1}^{L} P(q_{k'} \mid \beta_j^{(s)})^{y_{ij}} \left[1 - P(q_{k'} \mid \beta_j^{(s)})^{y_{ij}}\right]^{1-y_{ij}}} \quad .... (19)$$

أما في مرحلة عملية تعظيم توقع لوغاريتم الأرجحية، ولكي نقدر معالم الفقرة j نطبق

المعادلة الآتية:

$$\sum_{k=1}^{K} \frac{r_{jk}^{(s)} - n_k^{(s)} P(q_k \mid \beta_j)}{\left[1 - P(q_k \mid \beta_j)\right] \left(P(q_k \mid \beta_j)\right)} \frac{\partial P(q_k \mid \beta_j)}{\partial \beta_{ij}} = 0 \quad .... (20)$$

وتكرر العملية للوصول إلى ما يطلق عليه التقارب (convergence)، في نتائج التقدير

لمعالم الفقرة j والتي يرمز لها بالرمز Δ والقدرة π عند كل مستوى من مستويات القدرة q_k.

وقد أشار لورد في دراسة (lord,1986) إلى الدالة الرياضية للأرجحية العظمى في

استخراج معالم الفقرة كما يأتي:

$$Maximize \ \mathrm{L}(a,b,c) = \prod_{a=1}^{n} \int_{-\infty}^{\infty} g(\theta_a) L(\theta_a; a,b,c) d\theta_a \quad .... (21)$$

يشير هامبلتون وسوامينثان (Hambeleton & Swaminithan, 1985) و(عــلام، 2005) إلى إيجابيات وعيوب طريقة (MML)، ومن أهم ايجابياتها:

- تقدر معالم الفقرات للنماذج اللوغارتمية الأحادية والثنائية والثلاثية (1-PL,2-PL. 3-PL).
- تمتاز هذه الطريقة بفعاليتها (Efficiency) وباختلاف طول الاختبار.
- إعطاء قيم تقديرية لمعالم الفقرات الاختبارية، سواء كانت الإجابات صــحيحة أو خاطئــة لجميع هذه الفقرات، أي أن هذه الطريقة تحلل فقرات الاختبار جميعها دون حذف كما هو متبع في طريقة (JML) التي تشترط حذف الفقرات التي تكون جميــع الإجابــات عنهــا صحيحة أو خاطئة.

أما عيوب طريقة (MML):

- تحتاج هذه الطريقة إلى توزيع مستويات القدرة مسبقاً، قبل البدء بعملية تقــدير معــالم الفقرات، وهذا يجعل عملية التقدير معتمداً على ملاءمة مستويات القدرة الافتراضية.
- التقدير غير الدقيق لمعلمة التخمين بسبب ارتفاع قيمة الخطأ المعياري في التقدير، ممــا يؤثر سلبيا على تقدير معالم الفقرات الأخرى في النموذج اللوغاريتمي الثلاثي،(3-PL).
- تحتاج هذه الطريقة إلى حجم عينة كبير للوصول إلى دقة في تقدير معــالم الفقــرات. ويعود سبب زيادة حجم العينة من أجل الاقتراب من توزيعات مستويات القدرة، التــي نحتاجها لتقدير الأرجحية العظمى لمعالم الفقرات (a , b , c).

ب- طريقة بيز في تقدير معالم الفقرات (Bayesian Modal Estimation)

طريقة بيز (Bayes) في الاستدلال الإحصائي تعود للعالم البريطاني توماس بييــز (Thomas Bayes)، (Bellhouse, 2004) وتعتبر هذه الطريقة من الأســاليب الإحصــائية المستخدمة لتقدير معالم الفقرات في نماذج نظرية استجابة الفقرة، إضافة إلى طرق الأرجحيــة

العظمى سابقة الذكر (JML, CML, MML). وفي طريقة بييز يتم اعتبار المعلمة التي سيتم تقديرها بأنها متغير عشوائي تتبع توزيع احتمالي معين، أو ما يطلق عليه التوزيع القبلي (Prior Distribution)، وذلك من خلال الأخذ بعين الاعتبار المعلومات السابقة للمعلمة غير المعروفة في عملية التقدير، ويتم تحديد المعلومات السابقة عادة من قبل الباحث معتمداً على معتقد شخصي أي بناءً على خبرته السابقة، او من خلال الخصائص الإحصائية لذلك المعلم الذي سيتم تقديره (lord,1986) (Swaminithan & Gifford,1982).

وقد ذكر علام (2005) أن الأساليب التي تعتمد على نظرية بييز يمكن أن تساعد في الحصول على تقديرات جيدة. وأشار جو وجين (Gao &Chen,2005) إلى أن تقديرات بييز تعتمد على التوزيع السابق (التوزيع القبلي) أو ما يطلق عليه المعلومات الأولية، فإذا كان التوزيع السابق غير غني بالمعلومات فإن تقديرات بييز والأرجحية العظمى الهامشية تؤدي إلى النتائج ذاتها، أما إذا كان التوزيع السابق توزيعاً طبيعياً فإن طريقة بييز تكون أكثر دقة.

وقد أشار لورد (lord,1986) إلى معادلة بييز في تقدير معالم الفقرات كما يأتي:

$$Maximize\, f(\theta;a,b,c) = L(\theta;a,b,c)g_1(\theta)g_2(a,b,c) \quad ....(22)$$

$$L(\theta_j,a,b,c,=\prod_{a=1}^{N}\prod_{i=1}^{n} P_i\theta_a^{u_{ia}}Q_i\theta_a^{1-u_{ia}} \quad ....(23)$$

$P_i\theta_a^{u_{ia}}$: احتمال الإجابة الصحيحة (1)

$Q_i\theta_a$: احتمال الإجابة الخاطئة (0)

أما هامبلتون وسوامينثان (Hambeleton & Swaminithan, 1985) فقد عرضا معادلة بييز على النحو الآتي:

$$f(\theta,a,b,c|u)\infty L(u|\theta,a,b,c)\left\{\prod_{i=1}^{n} f(a_i)f(b_i)f(c_i)\right\}\prod_{a=1}^{N} f(\theta_a) \quad ....(24)$$

تشير ($f(\theta a)f(ai)f(bi)f(ci)$) إلى التوزيعات البعدية للمعالم، حيث أن:

قدرة الأفراد ($a\theta$) ($a = 1, \ldots\ldots\ldots, N$)

تمييز الفقرة (ai) ($i = 1, \ldots\ldots\ldots n$)

صعوبة الفقرة (bi) ($i = 1, \ldots\ldots\ldots n$)

تخمين الفقرة (ci) ($i = 1, \ldots\ldots\ldots n$)

ويمكن اضافة أو حذف المعلم بناءاً على النموذج اللوغاريتمي المستخدم، ويشترط تحديد توزيعات القدرة (θ) لإيجاد معالم الفقرة باستخدام طريقة بييز، ويتم إيجاد تقديرات قيم المعالم (θ, a, b, c) التي تعظم التوزيع اللاحق المشترك، وبذلك نحصل على تقدير لمعالم الفقرات.

ومما سبق يتضح بعض جوانب الاختلاف بين الدوال الرياضية لكل من أسلوبي الأرجحية العظمى وبييز في تقدير معالم الفقرات، ويكمن الفرق الجوهري بينهما أن أسلوب بييز يراعي التوزيعات السابقة للمعالم قبل تقديرها.

مشكلة الدراسة وأسئلتها:

في ظل وجود العديد من الطرق المستخدمة في تقدير معالم النماذج اللوغارتمية (الأحادي والثنائي والثلاثي)، ومنها الأرجحية العظمى وطريقة بييز يبقى التساؤل حول أكثر هذه الطرق فاعلية في دقة تقدير معالم الفقرة، وضمن شروط محددة، وبناء عليه فإن هذه الدراسة ستحاول فحص دقة تقديرات معالم الفقرة، في مواقف مختلفة وفق (عدد الفقرات، وعدد الافراد، والنماذج اللوغارتمية) لتحديد الأوضاع التي يفضل فيها استخدام طريقة بييز، والمواقف التي يفضل فيها استخدام طريقة الأرجحية العظمى، في تقدير المعالم وفق نماذج نظرية استجابة الفقرة، وعلى وجه التحديد ستحاول الدراسة الإجابة عن الاسئلة الآتية:

السؤال الأول:

"ما أثر طريقة التقدير (MML، وBayes) في دقة تقدير معالم الفقرة في ضوء اختلاف المتغيرات (النموذج اللوغاريتمي المستخدم وعدد الفقرات، وحجم العينة)؟"

السؤال الثاني:

"ما أثر طريقة التقدير (MML، وBayes) في الكفاءة النسبية لتقدير معالم الفقرة في ضوء اختلاف المتغيرات (النموذج اللوغاريتمي المستخدم وعدد الفقرات، وحجم العينة)؟"

السؤال الثالث:

"هل توجد فروق دالة إحصائيا عند مستوى الدلالة (α=0. 05) في قيم معاملات الارتباط بين معالم الفقرة الحقيقية وبين معالم الفقرة المقدرة وفق طريقتي التقدير (MML, Bayes) باختلاف المتغيرات (النموذج اللوغاريتمي المستخدم، عدد الفقرات، حجم العينة)؟"

هدف الدراسة:

طريقة تقدير معالم الفقرات والأفراد في نماذج نظرية استجابة الفقرة تؤثر على دقة التقدير، لذلك هدفت هذه الدراسة بشكل رئيس لتحديد أي من طرق التقدير (الأرجحية العظمى وبييز) أفضل في تقدير معالم الفقرة وتحت أي من المواقف المتعلقة باختلاف عدد الفقرات وحجم العينة، والنموذج اللوغاريتمي.

مبررات الدراسة:

الغاية الأساسية من نظريات القياس النفسي والتربوي (نظرية الاختبار ونظرية استجابة الفقرة) هو الوصول إلى مؤشرات عن دقة القياس، واستخدام الاختبارات الأكثر دقة في تقدير قيم القدرة للأفراد، وفي نظرية استجابة الفقرة تعتمد دقة تقدير معلمة القدرة على دقة تقدير

معالم الفقرات، لذلك يُسعى للوصول إلى طريقة لتقدير معالم الفقرات بأعلى درجة من الدقـــة، وكون هناك طريقتان لتقدير هذه المعالم (طريقة الأرجحية العظمى وطريقة بييز) جاءت هـــذه الدراسة للوقوف على أيٍّ من الطريقتين أكثر دقة في تقدير معالم الفقرات، ونظراً لعدم وجـــود إجماع حول أي الطريقتين أكثر دقة وكفاءة في تقدير معالم الفقرات، وتحت أي من المواقـــف (النموذج اللوغارتمي المستخدم، وعدد الفقرات، وحجم العينة)، كان القيام بهذه الدراسة مبرراً.

مصطلحات الدراسة :

1. دقة التقدير(Accuracy of Estimation):

هو تعبير يشير إلى جودة التقدير، من خلال اقتراب قيمة المعلم المقدر من القيمة الحقيقية، حيث يمكن الوصول إلى ذلك باختيار التقدير الذي يتصف تباينه بأنه أقل تبايناً مـــن أي تقـــدير آخر ويقاس باستخدام مربعات الأخطاء للتقدير، أو الخطأ المعياري فـــي التقـــدير، أو الفعاليـــة النسبية، أو عدم التحيز.

2. طريقة الأرجحية العظمى (Maximum Likelihood Estimation):

هي إحدى الطرق الإحصائية المستخدمة في إيجاد تقدير معالم الفقرة وهـــي التمييـــز والصعوبة والتخمين (a ,b ,c) ومعلمة القدرة (θ) في نظرية استجابة الفقـــرة، وتقـــوم هـــذه الطريقة على أساس إيجاد قيمة معالم الفقرة ومعلمة القدرة التي تعمـــل علـــى تعظـــيم اقتـــران الاحتمالية (Likelihood Function)

3. طريقة بييز (Bayesian estimation) :

هي إحدى الطرق الإحصائية المستخدمة في إيجاد تقـــدير معـــالم الفقـــرة: التمييـــز، والصعوبة، والتخمين (a ,b ,c) والقدرة (θ) في نظرية استجابة الفقرة، وتقوم هذه الطريقة

على افترا ض أن هناك توزيعا احتمالياً مسبقاً قد يسهم في تحقيق دقة أكبر في عمليـــة تقـــدير معالم الفقرة، وهى أكثر تعقيدا وصعوبة في إدراكها مقارنة مع طريقة الأرجحية العظمى.

4. الكفاءة النسبية (RE: Relative Efficiency):

هو حاصل قسمة تباين القيم المقدرة لمعالم الفقرة (الصعوبة، التمييز) بطريقة Bayes على تباين القيم المقدرة لمعالم الفقرة (الصعوبة، التمييز) بطريقة MML، فإذا زاد ناتج القسمة عن 1 تكون الكفاءة النسبية لصالح MML وإذا كان ناتج القسمة أقل من 1 صحيح تكون الكفاءة النسبية لصالح Bayes، والذي يمثل بالصيغة الرياضية الآتية:

$$RE_{a \text{ or } b} = \frac{V\ \hat{a \text{ or } b}_{Bayes}}{V\ \hat{a \text{ or } b}_{MML}} \quad .... \quad (25)$$

5. الإحصائي (RMSE):

هو الجذر التربيعي لمعدل مربعات الخطأ (Root Mean Square Error: RMSE)، بين القيمة المقدرة والقيمة الحقيقة، وهو مؤشر لدقة تقدير المعالم، حيث يتم إيجاد القيمة للفـــرق بين القيم المولدة (الحقيقية) والقيم المقدرة بأي طريقة من طرق التقدير، وفقا للمعادلة الآتية:

$$RMSE_{x} = \sqrt{\frac{\sum_{i=1}^{n} (\hat{X}_{esimated} - X_{true})^2}{n}} \quad .... \quad (26)$$

6. **حجم العينة:** تعني عدد المختبرين الذين سيطبق عليهم الاختبار، وفي هذه الدراســـة يطبـــق على أربع عينات بأحجام مختلفة.

7. **معالم الفقرات:** هي قيم احصائية يتم تقديرها باستخدام معادلات رياضية، وتشمل (الصعوبة، التمييز، التخمين).

8. **معلمة الصعوبة Threshold:** مستوى القدرة الذي يقابل احتمال (0.50) للإجابة عن الفقرة إجابة صحيحة عندما يكون التخمين مساوياً للصفر.

9. **معلمة التمييز Slope:** نسبة ميل منحنى خصائص الفقرة، الذي يقابل النقطة التي تكون عندها القدرة تساوي صعوبة الفقرة.

محددات الدراسة:

1. تتحدد نتائج هذه الدراسة بالأسلوب الذي يستخدمه الباحث والقائم على توليد البيانات باستخدام برنامج (WinGen v1. 4 .)
2. نظراً لتعدد طرق الأرجحية العظمى في التقدير (JML, CML, MML) ستقتصر هذه الدراسة على طريقة الأرجحية العظمى الهامشية (MML) والمستخدمة في برنامج (Bilog-Mg 3).
3. ستقتصر هذه الدراسة على النموذج اللوغاريتمي أحادي المعلمة (1-PL) والنموذج اللوغاريتمي ثنائي المعلمة (2-PL) من نماذج نظرية استجابة الفقرة.
4. سيتم تقدير معالم الفقرات وفق طريقة بييز باستخدام برنامج (WinBUGS).

الفصل الثاني

الدراسات السابقة

تهتم هذه الدراسة في دقة تقدير معالم الفقرات، وأي من طريقتي التقـدير (Bayes, MML) أكثر دقة وكفاءة في عملية تقدير معالم الفقرات، وطبيعة التأثيرات التي يحدثها حجـم العينة وعدد الفقرات على دقة التقدير، فقد أشار لورد (lord, 1980) إلى أثر تفاعل حجم العينة وعدد الفقرات على جودة تقدير معالم الفقرات. لذا يقوم الباحث بهذا الفصل، بعرض الدراسات المرتبطة بدقة وكفاءة التقدير مرتبة ترتيبًا زمنياً من الأقدم إلى الأحدث، وسيتم التعليق على هذه الدراسات لاحقًا.

قام ساومينثان وجيفورد (Swaminathan & Gifford,1982) بدراسة هدفت لمقارنة تقديرات بييز والأرجحية العظمى في تقدير معالم الفقرات والقدرة وفق نموذج راش الأحـادي، وذلك باستخدام البيانات المولدة بواسـطة برنـامج (DATGEN) لعينتـين مـن المتقـدمين (n=50,75) مستخدماً اختبار تم تقسيمه في ثلاثة أطوال على النحو التالي (n=50,100,150) وخلصت الدراسة إلى أن هناك فروقًا بين طريقتي التقدير، وخاصة عند حجوم عينات وعـدد فقرات صغير. وأظهرت تقديرات (بييز) تحسنًا في تقدير معالم الفقرات عندما كانـت أحجـام العينات صغيرة هذا من جهة، وكانت تقديرات بييز ذات كفاءة أعلـى عنـدما طبقـت علـى الاختبارات القصيرة مقارنة مع الاختبارات الطويلة، أي أن تقديرات المعالم كانت أفضل عنـد حجم عينة (n=50) مقارنة مع الاختبارات (n=150) من جهة أخرى، وتشير النتائج لمقارنـة طريقة بييز مع طريقة الأرجحية العظمى أن التقديرات لكل منها تعطي نتائج متقاربة للعينـات الكبيرة من المتقدمين للاختبار، وكذلك تكون التقديرات متقاربة عندما نقارن بين عدد كبير مـن

الفقرات وقد استدل على ذلك من خلال إيجاد معامل الارتباط بين القيم المقدرة وفق الطريقتين، فقد كان معامل الارتباط قوي بين الطريقتين (بييز، الأرجحية العظمى) للعينات الكبيرة من الأفراد، وعدد الفقرات الاختبارية الكبيرة.

قام لورد (Lord, 1986) بدراسة، لتوضيح طريقة الأرجحية العظمى وبييز في تقدير معالم الفقرات، وبين أنه عندما يكون لدينا (1000) أو (2000) من المتقدمين لاختبار مكون من (40) فقرة او أكثر، يكون لدينا دقة نظرية في التقديرات بواسطة (MML) مقارنة مع (JML)، أما في حال أن كانت هناك (10) او (15) فقرة لكل شخص تعطي طريقة (JML) تقديرات منحازة للمعالم، مما يؤدي لتقديرات خاطئة رغم أن عدد المتقدمين كبير، وأشار إلى أن تقديرات بييز المعتمدة على التقديرات السابقة تقلل من متوسط مربع الخطأ (MSE) للتقدير مقارنة مع قيمة الأرجحية العظمى ويعود سبب ذلك إلى معرفتنا المسبقة بمعالم الفقرات، من هنا تظهر أهمية طريقة بييز للاختبارات الواقعية حيث يمكن الحصول على التوزيعات السابقة لاختبارات تم تطبيقها بشكل متكرر على المختبرين سنة بعد سنة، حيث يصبح من الممكن استنتاج توزيعات سابقة لمعالم الفقرة والقدرة من نتائج سابقة وفي هذه الحالة يمكن ان تعطي إجراءات بييز تقديرات أفضل من طريقة الأرجحية العظمى.

أجرى غولدمان وراجو (Goldman & Raju ,1986) دراسة هدفت إلى مقارنة نموذجين من نماذج الفقرة هما: النموذج اللوغاريتمي أحادي المعلم وثنائي المعلم، من حيث اختلافهما في تقدير معالم الفقرات، وقدرات الأفراد، وأي النموذجين أكثر ملاءمة للعينات الصغيرة، وقسمت عينة الدراسة المكونة من (3000) مفحوص في ثلاث عينات جزئية هي (n= 250, 500,1000) من أجل معرفة أثر حجم العينة على تقديرات معالم الفقرات الصعوبة والتمييز، أما أداة الدراسة فقد كانت اختبار استعداد للعمل، مكون من (78) فقرة، أظهرت نتائج

الدراسة أن تقديرات النموذج اللوغاريتمي أحادي المعلم أكثر دقة في تقدير المعالم عندما يكون حجم العينة صغير (250) مفحوص، ومن جهة أخرى عندما تكون حجوم العينات كبيرة نسبيا قرابة (1000,500) مفحوص، تكون دقة تقدير المعالم أفضل للنموذج اللوغاريتمي ثنائي المعلم مقارنة مع النموذج اللوغاريتمي أحادي المعلم.

وفي دراسة قام بها بوزيهي (Pozehi,1990) هدفت إلى تحديد الطول الأمثل للاختبار الذي تعطى عنده جميع النماذج المدروسة أعلى درجة ثبات واتساق في قرارات الإتقان، وكذلك تحديد نموذج من نماذج الفقرة تحقق مطابقة أفضل للبيانات، وذلك مع حجم العينة الكلية والطول الكامل للاختبار وتعيين النموذج الذي يعطي أكبر دقة ممكنة في تقديرات الصعوبة والقدرة مع الحالات المختلفة للاختبار، ومع أحجام مختلفة للعينة، مستخدماً خمسة نماذج من نماذج نظرية استجابة الفقرة هي: النموذج اللوغاريتمي الأحادي والنموذج اللوغاريتمي الأحادي المعدل والنموذج اللوغاريتمي الثنائي والنموذج اللوغاريتمي الثنائي المعدل والنموذج اللوغاريتمي الثلاثي، وقد بلغ حجم العينة (2039) ممرضة قسمت في عينات حجم كل منها (250) او(500) ممرضة، طبق عليها اختبار في الصحة العقلية، مكون من (150) فقرة، وقد توصلت الدراسة أن النموذج اللوغاريتمي الأحادي والنموذج اللوغاريتمي الأحادي المعدل، كانا أكثر دقة في تقدير القدرة والصعوبة من النماذج الأخرى، ولا توجد فروق تذكر بين النموذج اللوغاريتمي أحادي المعلم والنموذج اللوغاريتمي احادي المعلم المعدل، اما في حال استخدام العينات الكبيرة يكون النموذج اللوغاريتمي ثلاثي المعلم أكثر ملاءمة.

قام كل من بارنيز وايز (Barnes & Wise,1991) بدراسة هدفت إلى مقارنة النموذج الأحادي المعدل مع النموذج اللوغاريتمي الأحادي والنموذج اللوغاريتمي الثلاثي، حينما تكون أحجام العينات صغيرة، مستخدما أسلوب المحاكاة لتوليد عينات أحجامها (50, 100,200)

مفحوصاً، واختبارين بأطوال (25) فقرة و(50) فقرة، ومن خلال معاملات الارتباط بين القيم المولدة والقيم المقدرة توصلت الدراسة إلى النتائج الآتية: أن صعوبة الفقرة المقدرة كانت ذات ارتباط مرتفع مع البيانات المولدة في جميع النماذج، وكانت أعلاها في النموذج اللوغاريتمي الأحادي المعدل، أما النموذج اللوغاريتمي الثلاثي فقد أظهرت النتائج أن قيم الارتباطات تزداد مع زيادة عدد الفقرات وحجم العينة. أي بزيادة حجم العينة وعدد الفقرات نحصل على تقديرات أفضل لمعالم الفقرات. وأظهرت النتائج أن أخطاء التقدير كانت أقل ما يمكن مع النموذج اللوغاريتمي المعدل، يليه النموذج اللوغاريتمي الثلاثي، وأخيراً النموذج اللوغاريتمي الأحادي، كما أظهر أن هذه الأخطاء تتناقص بشكل واضح مع زيادة حجم العينة.

أما نيرنج (Nerng, 1995) فقد قام بدراسة هدفت إلى مقارنة طريقة الأرجحية العظمى وطريقة بييز في تقدير معالم الفقرات، معتمداً على أسلوب المحاكاة، من خلال توليد بيانات لـ (1000) مفحوص. وقد استخدم الباحث معامل الارتباط لإظهار التقارب بين الطريقتين في التقدير (طريقة الأرجحية العظمى وطريقة بييز)، وأظهرت نتائج معامل الارتباط بين القيم المقدرة بطريقة الأرجحية العظمى والقيم المقدرة بطريقة بييز أن هناك ارتباطًا قويًا بينهما، مما يشير إلى تكافؤ طريقتي التقدير.

قام ليندن (linden, 1998) بدراسة أسلوب بييز في إنتقاء الفقرات للاختبارات المكيفة، هدفت إلى فحص طرق تقدير معالم الفقرات (الصعوبة والتمييز) وفق النموذج اللوغاريتمي الثنائي، معتمداً عدة طرق لتحديد معايير انتقاء الفقرات للاختبارات التكيفية التي تعتمد على التوزيع البعدي الحقيقي. (true posteriors) وتألفت الدراسة من أربعة اختبارات، بأطوال مختلفة (5,10,20,30) فقرة، وأظهرت النتائج أن طريقة بييز أكثر دقة في تقدير معالم الفقرات

الصعوبة والتمييز عند استخدام عدد الفقرات القليل، مقارنة مع طريقة الأرجحية العظمى معتمداً على المؤشر الإحصائي متوسط مربعات الأخطاء ((Mean Squared Error (MSE)).

وأما دراسة عليان والدرابيع (2000)، فقد هدفت هذه الدراسة التي استخدمت أسلوب المحاكاة في توليد البيانات، إلى دراسة أثــر حجم العينة وطول الاختبار والتفاعل على دقة تقدير قدرة الفرد ومعالم الفقــرة، عنــد اســتخدام النموذج اللوغاريتمي ذي المعالم الثلاثة، وتم توليد البيانـــات بواقع 30 مرة داخل كل خلية من خلايا التصميم، ثم التحقق من دقة التقدير لقدرة الفرد ومعالم الفقرة باستخدام معامل الارتباط بين القيم الحقيقية لهذه المعالم والقيم المقدرة. حيث تم توليــد بيانـــات قدرة الفرد التي تتراوح بين (+3، -3) من توزيع طبيعي متوسطه صفر وانحرافـــه المعيـــاري واحد، وذلك تم توليد قيم معاملات صعوبة وتمييز الفقرات من توزيـــع طبيعـــي متوســـطه صـــفر وانحرافه المعياري واحد لقيم صعوبة الفقرات، ومن مدى تراوح بـــين (2 إلى 0.5) بوســـط حـــسابي مقداره (1.25) وانحراف معياري مقداره (0.25) لقيم تمييز الفقرات، بذلك تم توليد قـــيم معـــاملات التخمين من توزيع أحادي بمدى تراوح بين (0.1 إلى 0.3) وبوسط مقداره (0.20) وانحـــراف معيـــاري مقداره (0.5)، وللإجابة عن فرضيات الدراسة، تم استخدام تحليل التباين الثنائي، حيـــث أشـــارت النتائج إلى وجود فرق معنوي لتفاعل متغيري حجم العينة وطول الاختبار على دقة تقدير معلمة صعوبة الفقرة، وأظهرت النتائج أن دقة التقدير لصعوبة الفقرة هي أفضل ما يمكـــن عنـــدما يتكـــون الاختبار من 80 فقرة، وحجم العينة من 320 فردا، كذلك أظهرت النتائج أن متغير حجم العينـــة له أثر معنوي عند تقدير معالم الفقرة، ولذلك عند استخدام حجم كبير للعينة، وأخيـــراً أظهـــرت النتائج وجود أثر معنوي لمتغير طول الاختبار، وذلك عند اســـتخدام عـــدد كبيـــر مـــن فقـــرات الاختبار.

كما قام الدرابيع (2001) بدراسة هدفت إلى معرفة فعالية النمـوذج اللوغـاريتمي ذي المعلمة الواحدة – نموذج راش – في تقدير قدرة الأفراد ومعامل صعوبة الفقرة باختلاف حجـم العينة وطول الاختبار، واستخدم الباحث برنامج (IRTDATA) لتوليد بيانات ثنائية الاستجابة، لثلاث اختبارات بأطوال مختلفة تبلغ (300,50,25) فقرة، لعدد من عينات الأفـراد بأحجـام مختلفة تبلغ (50, 100, 500=n) وبعد ذلك مررت البيانات المولدة لبرنامج (Bilog-Mg 3) بهدف تقدير معالم الفقرة والافراد، وأظهـرت نتـائج تحليـل التبـاين الثنـائي (two way ANOVA) وجود فروق ذات دلالة احصائية لتفاعل طول الاختبار وحجم العينة مع دقة تقدير معلمة الفقرة وقدرات الأفرد.

أجرى ديمارس (Demars, 2001) دراسة هدفت إلى مقارنة أثر حجم العينـة علـى النموذج اللوغاريتمي الأحادي والنموذج اللوغاريتمي الثلاثي في تقديرات معالم الفقرة، مستخدماً بيانات مولدة لـ (2000) طالب، يخضعون لاختبارات مختلفة في فقراتها، تراوح عدد فقراتها بين (15) و(60) فقرة، وكانت معالم الفقرات متساوية في وسطها وانحرافها المعياري لقدرات الأفراد في المجموعات. استخدم الباحث برنامج (Bilog-Mg 3) وفـق طريقـة الأرجحيـة العظمى الهامشية، لتقدير معالم الفقرات، ومن نتائج هذه الدراسة أن أثر حجم العينة كان متكافئًا في تأثيره في النموذج اللوغاريتمي أحادي المعلمة والنموذج اللوغاريتمي ثلاثي المعلمـة، وأن تناقص حجم العينة يؤثر بشكل سلبي في دقة تقدير معالم الفقرات وفق النمـوذج اللوغـاريتمي الأحادي والثلاثي.

قام فيتزباترك وآن (Fitzpatrick, & Ann, 2001) بدراسة هدفت إلى فحـص أثـر طول الاختبار وحجم العينة على ثبات الاختبار، مستخدماً أسلوب المحاكاة لتوليـد اسـتجابات لمفحوصين وفق العينات الآتية (200,500,1000) تخضع لاختبارات بأطوال مختلفة (2) فقرة

و (4) فقرات و(8) فقرات و(12) فقرة، لتقييم أثر طول الاختبار وحجم العينة على دقة تقـدير المعالم، واستخدم الباحث إحصائي الجذر التربيعي لمعدل مربعات الفروق (RMSD) كمؤشـر لدقة تقدير المعالم، وقد توصلت الدراسة إلى عدة نتائج منها أن دقة تقـدير المعـالم تـنخفض بانخفاض حجم العينة، فكانت تقدير المعالم أكثر دقة عندما كان حجم العينة كبيراً (1000) فرداً، وجاءت العينة التي تبلغ (500) فرداً في المرتبة الثانية، أما العينة التي تبلغ (200) فرداً فقـد أظهرت دقةً أقل وفق معيار (RMSD) مما يشير إلى إن زيادة حجم العينة يزيد من دقة تقدير المعالم.

قام كل من هيونغ ولوهسس ولين وشـين (Huang , Lohss, Lin & Shin, 2001) بدراسة لمقارنة فعالية بعض البرامج وهي (Bilog, Bilog – Mg, pic) في معايرة فقـرات اختبار إجازة، باختلاف حجم العينة، ولفحص أثر حجم العينة في دقة تقدير معالم الفقرات تـم استخدام النموذج اللوغاريتمي ثلاثي المعلمة بمعايرة (360) فقرة ثنائيـة التـدريج مسـتخدماً طريقة المحاكاة للحصول على بيانات مولدة، حيث تم افتراض أن أول (120) فقرة مولدة تمثل الجزء المشترك من الاختبار، أما الفقرات المتبقية (240) فهي مقسمة فـي أربعـة مسـتويات (اختبارات) يتضمن كلٌ منها مجال الاختصاص، وتم اختيار حجم عينة بمدى يتراوح بين (250) مفحوص و (1500) مفحوص، تحت افتراض أن تقديرات المفحوصين تتوزع توزيعاً طبيعيـاً وتم تشكيل (14) حالة حسب حجم العينة وتوزيعات القدرة، ولكل موقف من مواقف الدراسة تم معايرة الاختبار المكون من (360) فقرة، وذلك بتمرير الاختبار على البرامج الآتية (Bilog, Bilog – Mg, pic) لمعايرتها، واستخدم الباحث إحصائي الجذر التربيعي لمعـدل مربعـات الفروق (RMSD) لمقارنة تقديرات المعالم وفق الطرق المسـتخدمة، وأظهـرت النتـائج أن

أخطاء التقدير لمعلم الصعوبة ومعلم التمييز تكون أكبر عندما يقل حجم العينة في حال استخدام برنامج (Bilog).

أجرى العبابنة (2004) دراسة لمعرفة أثر حجم العينة، وطريقة انتقائها، وعدد الفقرات وطريقة انتقائها في دقة التقديرات لمعالم الفقرات والقدرات، حيث قام الباحث ببناء اختبار قدرة عقلية مؤلف من أربعة اختبارات فرعية هي (اختبار المفردات، اختبار المتشـابهات، واختبـار المتضادات، واختبار الحساب) وتكون الاختبار في صورته النهائية مـن (70) فقـرة، وطبـق الاختبار على عينة من طلبة الصف السابع الاساسي ذكوراً واناثاً في منطقة اربد الاولى، وقـد بلغ عدد أفراد العينة (1000) طالب وطالبة واقتصرت هذه الدراسة على أسـلوب الأرجحيـة العظمى الهامشية (MML) في برنامج (BILOG MG-3) لتقدير قدرة المفحوصين ومعـالم الفقرات على النموذج اللوغاريتمي الثلاثي. وتوصل الباحث إلى النتائج الآتية: تزداد دقة التقدير بزيادة حجم العينة وعدد الفقرات، وتتسم تقديرات معالم الفقرة بعدم الاستقرار عنـد اسـتخدام عينات مفحوصين مختلفة القدرة وتتسم تقديرات معلمة الصعوبة والقدرة بالدقة عنـدما يتوافـق مدى قدرة المفحوصين مع مدى صعوبة الفقرة، أما تقدير معلمة التمييز فتزداد بزيادة تباين قدرة المفحوصين، وأظهرت النتائج أيضاً أن دقة تقدير معلمة التخمين تزداد عند استخدام عينة مـن ذوي القدرة المتدنية.

قام كل من جيو وشين (Geo & chen, 2005) بدراسة لمقارنـة تقـديرات معـالم الفقرات (الصعوبة، التمييز، التخمين) التي تم الحصول عليهـا بطريقـة الأرجحيـة العظمـى الهامشية وطريقة بييـز حجـم العينـة (n=100,500,2000) فـرداً، وأطـوال اختبـارات (N=10,30,60) فقرة ثنائية الاجابة، وبعد أن تم توليد البيانات وفق الشـروط السـابقة علـى النموذج اللوغاريتمي الثلاثي (3-PL) تم مقارنة تقدير معالم الفقرات وفق الأسلوبين (Bayes,

(MML علماً أن الباحث قد استخدم برنامج (ETIRM) لتقدير معالم الفقرات وفق طريقة بييز، أما معالم الفقرات المقدرة وفق طريقة الأرجحية العظمى الهامشية قدرت بواسطة برنامج (Bilog-MG3). وطبقت الاختبارات الثلاث على عينة المختبرين (n=100) ثم طبقت الاختبارات على عينة المختبرين (n=500) واخيراً طبق الاختبار بأطواله الثلاثة على عينة المختبرين (n=2000). وأشارت نتائج الدراسة إلى أن تقديرات طريقة بييز أكثر دقة من تقديرات طريقة الأرجحية العظمى وذلك استناداً على قيم معاملات الارتباط، بين القيم المقدرة والقيم الحقيقية عندما يكون حجم العينة صغيرة نسبيًا (أقل من 500)، أظهر التوزيع السابق أثرًا كبيرًا على نتائج طريقة بييز، أما في حالة أن تكون العينات كبيرة بحجم (2000) تُظهر التوزيعات السابقة أثراً معتدلاً جداً (أو حتى غير ملحوظ على تقديرات المعالم)، وبشكل عام كانت تقديرات الأرجحية العظمى الهامشية وبييز تميل إلى إعطاء نتائج متشابهة ودقيقة عندما يكون حجم العينة كبيراً، وعندما يصبح حجم العينة كبيراً يكون أداء الإجرائين جيداً وغالباً ما يعطيان نتائج لا يمكن فصلهما حول المعالم الثلاثة في النموذج اللوغاريتمي الثلاثي (3-PL).

وفي دراسة قام بها غاري وفيرمونت (Garee & Vermunt ,2006) هدفت إلى تفادي تقديرات أطراف المتصل في نظرية استجابة الفقرة باستخدام التوزيع القبلي البييزي، اعتمدت هذه الدراسة البيانات الحقيقية لاختبارين، الاختبار الأول بعدد (5) فقرات والاختبار الثاني(9) فقرات وبحجم عينة (1000, 100=n)، وقد استخُدم في هذه الدراسةً بالإضافة لطريقة بييز طريقة الأرجحية العظمى لتقدير معالم الفقرات، وأظهرت النتائج أن تقديرات بييز أكثر ثباتاً في تقدير المعالم مقارنة بطريقة الأرجحية العظمى، وخاصة عند تقدير القدرات لأطراف متصل القدرة (أي أصحاب القدرات المرتفعة، وأصحاب القدرات المنخفضة).

قام العبابنة (2007) بدراسة هدفت إلى مقارنة فاعلية طريقة الأرجحية العظمى وطريقة بييز في تقدير معلمة القدرة عند استخدام النموذج اللوغاريتمي ثلاثي المعلمة، وقد استخدم الباحث اختبار قدرة عقلية، قد طوره في دراسة سابقة، مكون من (70) فقرة طبق على (1000) مفحوص وقد قام بمعايرة فقرات الاختبار وقدرات الأفراد باستخدام برنامج -Bilog) (MG 3 وفق النموذج اللوغاريتمي الثلاثي للحصول على قدرات الأفراد وفق طريقة الأرجحية العظمى وطريقة بييز. وأظهرت النتائج أن طريقة بييز تعمل بصورة أفضل من طريقة الأرجحية العظمى عند معايرة الفقرات بعينة ذوي القدرة العالية وعينة الأفراد ذوي القدرة المتدنية، وهذا يعني أن طريقة بييز تقدم تقديرات أكثر دقة من طريقة الأرجحية العظمى، بالاعتماد على مؤشر الكفاءة النسبية فيما أظهرت طريقة بييز تفوقا على طريقة الأرجحية العظمى عند المعايرة بالعينة العشوائية، وهذا يشير إلى أن طريقة بييز تكون مناسبة لتقدير القدرات عند أطراف متصل القدرة، وبينت نتائج الدراسة أن طريقة الأرجحية العظمى تعمل بصورة أفضل من طريقة بييز عند استخدام جميع فقرات المقياس، وتشير النتائج إلى أن طريقة بييز أكثر دقة في التقدير عند مستويات القدرة المتدنية، وعندما ترتفع قدرة المفحوصين لتصل إلى مستويات مقاربة لصعوبة الفقرة، وفي الاختبار الصعب تعمل طريقة الأرجحية العظمى بصورة جيدة مقارنة مع طريقة بييز.

أجرى كل من ليو، شولز ويو (Liu, Schulz & Yu, 2009)، دراسة هدفت إلى تقييم استخدام طريقة سلسلة ماركوف مونتي كارلو (MCMC) كخيار اضافي او بديل لتقديرات دقة مساواة العلامة الحقيقية للنموذج اللوغاريتمي الثلاثي (3-PL)، فقد تم فحص تقديرات الخطأ المعياري للتقدير لمعالم الفقرات في سياق علامات ثلاث حالات للاختبار هي: المتوازي، ومكافئ T، والمتجانس، واستخدمت البيانات المقلدة (المحاكاة) في هذه الدراسة، من خلال توليد

عينة الدراسة باستخدام برنامج (SAS) وبعد ذلك ادخلت البيانات لبرنـامجي (WinBUGS, Bilog)، علماً أن برنامج (WinBUGS) يستخدم طريقة (MCMC) في معايرة المعالم وفق طريقة بييز. وأظهرت النتائج في حالة الاختبارات المتكافئـة (T) تكـون تقـديرات الخطـأ المعياري الناتجة من طريقه (MCMC) كانت أكبر نسبياً من تلـك الناتجـة عـن الطريقـة المستخدمة في برنامج (Bilog) .

قام الثوابية (2010) بدراسة هدفت إلى استقصاء أثر حجم العينة فـي تقـدير معلمـة صعوبة الفقرات والخطأ المعياري في تقديرها، ولتحقيق ذلك تم إعداد اختبار رياضيات للصف العاشر الاساسي وكان عدد فقراته (80)، وعدد الطلبة الذين طبق عليهم الاختبـار (11292)، وتم تقسيم الطلبة في عدد من العينات، وتم استخدام طريقـة الأرجحيـة العظمـى الهامشـية (MML) من خلال برنامج (Bilog-MG3) لتقدير معلمة الصعوبة والخطأ المعيـاري فـي التقدير، وقد توصلت الدراسة إلى أن صعوبة الفقرة تزداد بزيادة حجم العينة، فقد كان متوسـط صعوبة الفقرات (.31) لوجيت عندما كان حجم العينة يساوي (200) وبسبب زيادة حجم العينة أرتفع متوسط صعوبة الفقرة إلى (1.1) لوجيت وذلك عند حجم عينة يبلـغ (11292) طالـب وطالبة. أما الخطأ المعياري في التقدير فقد تناقص بزيادة عدد أفراد العينـة، فبلـغ متوسـط الأخطاء المعيارية في التقدير (.32) لوجيت عندما كان حجم العينة (200)، وعند زيادة حجـم العينة (11292) أصبح الخطأ المعياري في التقدير (0. 07) لوجيت.

وأجرى دي لا توري ويوان (De La Torre & Yuan, 2010) دراسة هـدفت إلـى التعرف على أثر حجم العينة في دقة تقدير معلمة القدرة ومعالم الفقرة في اختبارات معدة حسب نماذج نظرية استجابة الفقرة، وتم اعتماد بيانات مولدة وفق مواقف مختلفة، من أجل التعرف إلى العلاقات بين تقديرات معلمة القدرة ومعالم الفقرة في الاختبار، وأثر حجم العينة في دقة تقـدير

معلمة القدرة ومعالم الفقرة، وأشارت النتائج إلى أن حجم العينة يؤثر في تقدير معلمة القدرة ومعالم الفقرة.

قام بيرجوس (Burgos, 2010) بدراسة هدفت إلى تقديم طريقة بييز في مجال نظرية استجابة الفقرة (IRT)، وقد طبق الباحث دراسته على نموذج راش أحادي المعلمة من خلال توليد بيانات للأفراد (n=500) واختبار بعدد فقرات (N=11) وتم التوصل إلى نتائج مفادها أن طريقة بييز تعطي تقديرات لمعالم الفقرات والأفراد أكثر ملاءمة وإقناعاً لتقدير نماذج (IRT) من طرق (MML,CML,JML) وأشار الباحث إلى أن طرق الأرجحية العظمى بدأت تصبح أقل جدوى حسابية، لذلك تم التوسع بها لاستيعاب الاثار العشوائية الكثيرة على دقة التقدير لذلك تعددت الطرق للأرجحية العظمى، فقد أظهرت الدراسة من خلال المقارنة بين طريقتي التقدير(الأرجحية العظمى وبييز) أن الفرق بين الدرجة الحقيقية والدرجة المقدرة بطريقة بييز أقل او تساوي طريقة الأرجحية العظمى مما يعطي الأفضلية لطريقة بييز في تقدير معالم الفقرة.

قام هسيه وآخرون (Hsieh, Proctor, Hou & Teo, 2010) بدراسة تهدف إلى مقارنة طريقة بييز مع طريقة الأرجحية العظمى في تقدير معالم الفقرات للنموذج اللوغاريتمي ثنائي المعلمة، مستخدما البيانات المولدة، حيث تم الحصول على حجم عينة (n=1000) وعدد فقرات (N=30) فقرة ثنائية الإجابة الصحيحة تأخذ (1)، والإجابة الخاطئة تأخذ (0)، وقدرت المعالم (الصعوبة، التمييز) وفق طريقتي الأرجحية العظمى الهامشية (MML) باستخدام برنامج (Bilog). وتقديرات طريقة بييز(Bayes) باستخدام برنامج (Winbigs). وأظهرت النتائج افضلية طريقة بييز مقارنة بطريقة الأرجحية العظمى الهامشية.

قام الشريفين (2012) بدراسة هدفت إلى الكشف عن أثر طريقة تقدير معالم الفقرات وقدرات الأفراد على قيم معالم الفقرة، والخصائص السيكومترية للاختبار في ضوء تغير حجم

العينة، واعتمدت الدراسة على بيانات حقيقية لاختبار فيزياء مكون من (23) فقرة مـــن اربـــع بدائل وطبق على (1000) طالب وطالبة، وتم تحليل النتائج وفق النمـــوذج الثلاثـــي (3-PL) باختلاف طريقة التقدير، مستخدما برنامج Bilog MG-3 في تقدير المعالم، وللوصول للنتائج تم استخدام الإحصاء الوصفي وتحليل تباين القياسات المتكررة، وأشارت النتائج إلـــى وجـــود فروق دالة إحصائيًا لمتوسطات الأخطاء المعيارية لتقديرات قدرات الأفراد، تعزى للتفاعل بين طريقة التقدير وحجم العينة، لصالح طريقة بيز في التقدير وخاصة عند حجوم العينات الصغيرة في حين لم تظهر فروق ذات دلالة إحصائيةً تعزى لمتغير حجم العينة أو طريقة التقدير.

تعقيب على الدراسات السابقة

عند استعراض الدراسات السابقة حول دقة التقدير، نجد أن الغالبية العظمى من الدراسات كانت مهتمة في أثر حجم العينة وطول الاختبار على دقة التقدير وفق طريقة الأرجحية العظمى، ويعزى ذلك إلى أن معظم البرمجيات المتاحة تقدر معالم الفقرات وفق هذه الطريقـــة، ونـــدرة البرمجيات الحاسوبية التي تقدر معالم الفقرات وفق طريقة (Bayes). ويمكن استخلاص بعض الاتجاهات حول النتائج التي تم التوصل اليها وفق هذه الدراسات:

1. معظم الدراسات بحثت دقة التقدير لمعالم الفقرات وفق طريقة الأرجحيـــة العظمـــى فقـــط المستخدمة في معظم البرامج الحاســـوبية مثـــل (Bilog mg3 , Multilog , Parscale ,TestFACT)
2. التركيز على حجم العينة وعدد الفقرات وأثرهما على دقة التقدير
3. معظم الدراسات التي قارنت بين دقة تقدير نماذج نظرية استجابة الفقرة، استخدمت أسلوب المحاكاة (البيانات المولدة)؛ حتى تحقق أقصى درجات الضبط التجريبي لأن البيانـــات تـــم أختيارها بطريقة عشوائية.

4. أظهرت الدراسات طرقًا مختلفة في الحكم على دقة التقـدير بعضـــها أســـتخدم معـــاملات الارتباط وأخرى استندت على مؤشر التحيز والبعض الاخر أستخدم مؤشر (RMSD).
5. أظهرت نتائج الدراسات تفاوت في الحكم على دقة التقدير، وفق طريقتي التقدير (الأرجحية العظمى وبييز) منها من أشار إلى أفضلية طريقة بييز في التقدير مثال دراســة بيرجــوس (Burgos, 2010) ودراسة هسيه وآخــرون (Hsieh, Proctor, Hou & Teo, 2010) ومنها ما أظهرت نتائجها أفضلية طريقة الأرجحية العظمى مثل دراسة كلٍ من ليو، شــولز ويو (Liu, Schulz & Yu, 2009)، وأخرى أشارت إلى تكافوء الطريقتين مثل دراســة نيرنج (Nerng, 1995) ودراسة جيو وشين (Geo & chen, 2005).
6. لم تقم أي دراسة عربية – في حدود علم الباحث- بتقدير معالم الفقرات وفق منحى بييــز (Bayes) الاستدلالي للوقوف على دقة تقدير معالم الفقرات مقارنةً مع الطرق الأخرى في التقدير.
7. نقص الدراسات وخاصة العربية منها التي تقدر معالم الفقرات وفق منحى بييز الاستدلالي؛ ويعود سبب ذلك إلى ندرة البرامج الحاسوبية التي تعمل وفق طريقة بييز الاحتمالية

الفصل الثالث

الطريقة والإجراءات

يتناول هذا الفصل وصفا دقيقاً للبيانات التي تم توليدها باستخدام برنامج (WinGen V. 3) والطريقة التي تم بها تقدير معالم الفقرات، الصعوبة (b) في النموذج اللوغاريتمي الأحادي، والصعوبة والتمييز (a, b) في النموذج اللوغاريتمي الثنائي، وفق طريقتي التقدير: طريقة الأرجحية العظمى الهامشية (MML) من خلال برنامج (Bilog- mg 3)، وطريقة بييز (Bayes) من خلال برنامج (WinBUGS V 1. 4)، واستخدم الباحث في هذه الدراسة المنهج التجريبي وذلك من أجل الإجابة عن أسئلة الدراسة المتعلقة بدقة وكفاءة تقدير معالم الفقرات وفقاً لطريقة (MML) وطريقة (Bayes) باختلاف النموذج اللوغاريتمي وعدد الفقرات وحجم العينة.

توليد البيانات (Simulating The data):

وفي هذه الدراسة استخدمت طريقة المحاكاة لتوليد البيانات اللازمة للبحث، المسماة طريقة مونت كارلو Monte Carlo Methods (MCM). وطبقت طريقة المونتي كارلو (MC) في مجال نظرية استجابة الفقرة IRT ؛ لتقدم معلومات حول دقة نماذج استجابة الفقرة في تقدير المعالم، وذلك من خلال البرامج الحاسوبية المتاحة لتوليد استجابات الأفراد على الفقرة، وتقدير معالم الفقرة، وقد استخدام برنامج (WinGen V. 3) لتوليد البيانات، وقد أكد العديد من الباحثين مثل (Christion & George,2009) على أن التجربة من خلال (MCM) تقدم نتائج التجربة الحقيقية نفسها التي طبقت على الأفراد، ويمتاز برنامج (WinGen V. 3) بمجموعة من المزايا التي جعلت الباحث يعتمد علية في توليد بيانات البحث منها:

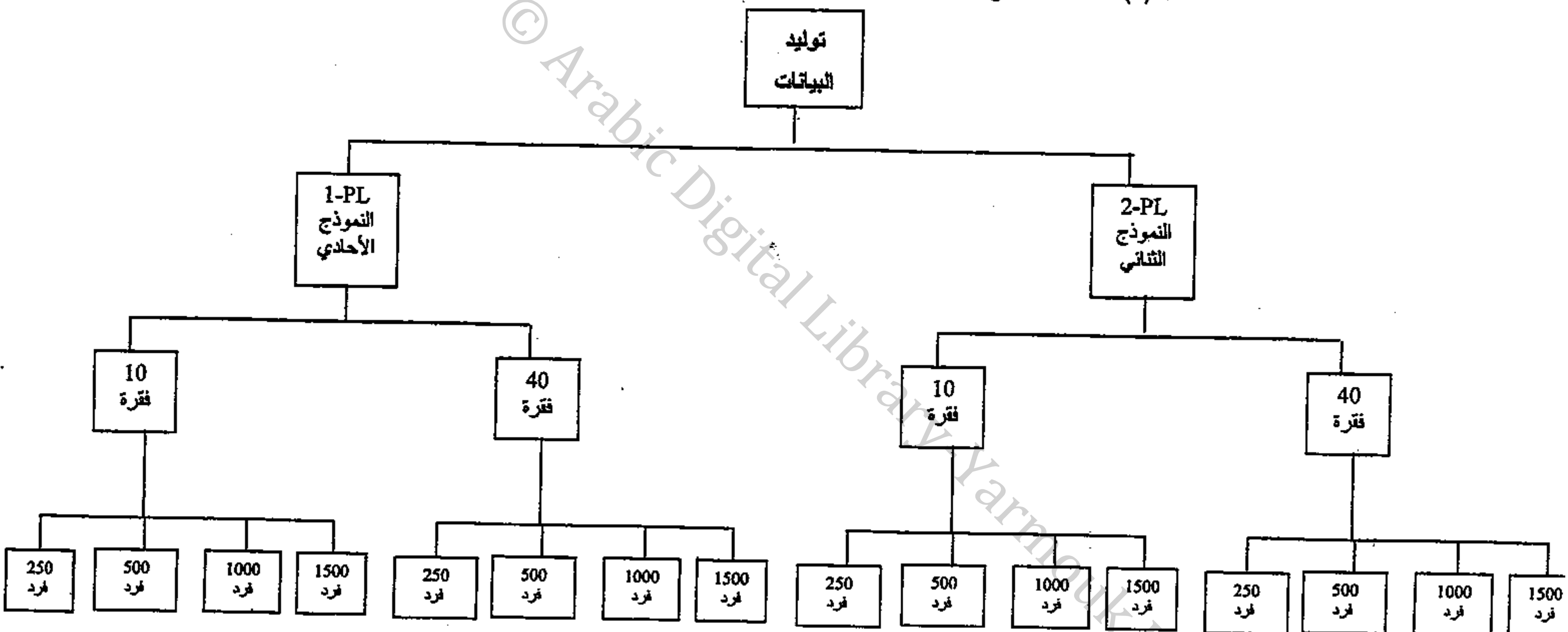
1. برنامج (WinGen V. 3) معد للاستخدام مع نماذج مختلفة لنظرية استجابة الفقرة، التـي تستخدم بشكل واسع، مثل نماذج نظرية استجابة الفقرة ثنائية الاسـتجابة Dichotomous IRT models (الأحادي والثنائي والثلاثي)، ونماذج نظريـة اسـتجابة الفقـرة متعـددة الاستجابة Polytomous IRT models، والنمـاذج اللابارامتريـة non parametric models.
2. برنامج (WinGen-3) يستطيع توليد معالم نظرية استجابة الفقرة من توزيعـات مختلفـة للمقاربة مع البيانات الواقعية، إذ يمكن للباحث من خلال هذا البرنـامج اختيـار التوزيـع الطبيعي، أو المنتظم، أو الطبيعي اللوغاريتمي، أو توزيع بيتا الذي يعتمد علـى توزيعـات ملتوية، كما يمكن الباحث من التحكم في بارامترات الفقرة والفرد، من خلال السماح بإدخال قيم هذه البارامترات أو استخدام الأرقام العشوائية التي يرغب بها عند توليـد بـارامترات النموذج.
3. برنامج (WinGen-3) سهل الاستخدام ويملك طرق تشغيل متقدمة من خلال نافذة ويندوز (Windows) كما أنه يمتلك حلولاً متقدمة لمحاكات البيانات الملائمـة لنمـاذج نظريـة استجابة الفقرة، وتمر عملية توليد البيانات في هذا البرنامج في ثلاث مراحل وهي (توليـد بارامترات الفرد، توليد بارامترات الفقرة، محاكات بيانات استجابة الفقرة)، وجميع المراحل الثلاثة يمكن مشاهدتها على شاشة واحدة، كما أن برنامج (WinGen-3) يـزود الباحـث برسومات بيانية لمنحنيات متعددة مثل المنحى المميز للفقـرة (ICC)، ومنحنيـات دالـة معلومات الفقرة (IFC)، ومنحنيات دالة معلومات الاختبار (TCC).
4. برنامج (WinGen V. 3) أداة بحث مجهزة لأهداف الدراسات المختلفة، فمـثلاً بيانـات الإعادة (Replication) يمكن محاكاتها بحدود (1000000) مجموعة، كمـا أن ملفـات

(Syntax) يتم تزويدها لمعظم برامج تحليلات نظرية استجابة الفقرة مثل برنـــامج Bilog (Zimowski, Muraki, Mislevy & Bock, 2003)، أما برنامج .WinBUGS v1) 4) يتطلب تغيير الرابط لملف بيانات برنامج (WinGen v. 3) حتى يتم استقباله.

خطوات المحاكاة في توليد البيانات:

قام الباحث بتوليد البيانات باستخدام برنامج (WinGen v. 3) وفق متغيرات الدراســـة (النموذج اللوغاريتمي، عدد الفقرات، حجم العينة)، المبينة في الشكل (1)، حيث كـــان هنالـــك (16) موقفاً من أجل المقارنة بين دقة وكفاءة تقدير معالم الفقرة باســـتخدام طريقـــة (MML) والمقدرة بواسطة برنامج (Bilog-MG 3) وطريقـــة (Bayes) المقـــدرة بواســـطة برنـــامج (WinBUGS V 1.4). علماً أن البيانات المولدة للمعالم تم الحصول عليها باستخدام برنـــامج (WinGen v. 3)، تحت افتراض التوزيع الطبيعي لمعلمة القدرة (θ) بمتوســـط حســـابي (0) وانحراف معياري (1) والتوزيع المنتظم لمعلمة التمييز بقيمة ابتدائيـــة (0.4) وقيمـــة نهائيـــة (1.7)، أما معلمة الصعوبة فقد تم توليدها تحت افتراض التوزيع الطبيعي بمتوسط حسابي (0) وانحراف معياري (1) للنموذج اللوغاريتمي أحادي المعلمة (1-PL)، ثم النموذج اللوغاريتمي ثنـــــائي المعلمــــة (2-PL)، عنـــــد حجـــــم العينــــة (1500,1000,500,250) وعدد فقرات (40,10).

الشكل (1) مخطط للنماذج اللوغاريتمية وحجوم العينات وعدد الفقرات المستخدمة في الدراسة



والجدول رقم (1) يظهر قيم الوسط الحسابي والانحراف المعياري لقيم معالم الفقرات المولدة (الحقيقية) التي تم الحصول عليها من برنامج (WinGen v. 3) وفق النموذج اللوغاريتمي (1-PL , 2-PL) وعدد الفقرات.

الجدول (1)

الإحصاء الوصفي لمعالم الفقرات المولدة (الحقيقية) وفقا للنموذج اللوغاريتمي وعدد الفقرات

النموذج	عدد الفقرات	المعلمة	الوسط الحسابي	الانحراف المعياري
أحادي المعلم	10	معلمة الصعوبة الحقيقية	0.216	1.040
	40	معلمة الصعوبة الحقيقية	-0.161	1.060
ثنائي المعلم	10	معلمة التمييز الحقيقية	1.111	0.194
		معلمة الصعوبة الحقيقية	0.216	1.040
	40	معلمة التمييز الحقيقية	1.035	0.248
		معلمة الصعوبة الحقيقية	-0.161	1.060

تم التحقق من أحادية البعد للبيانات، بالاعتماد على نسبة التباين المفسر، والجذر الكامن وScree Plot، ويظهر الجدول (2) نتائج التحليل العاملي للبيانات المولدة باختلاف النموذج اللوغاريتمي (1-PL, 2-PL) وحجم العينة (1500,1000,500,250) وعدد الفقرات (10, 40) بهدف التحقق من افتراض أحادية البعد.

الجدول (2)

نتائج التحليل العاملي للبيانات المولدة باختلاف النموذج اللوغاريتمي (1-PL, 2-PL) وحجم العينة (1500,1000,500,250) وعدد الفقرات(40,10)

طول الاختبار	حجم العينة	المكون	النموذج: أحادي المعلمة: الجذر الكامن	التباين المفسر	التباين المفسر التراكمي	مؤشر (2)	مؤشر (3)	ثنائي المعلمة: الجذر الكامن	التباين المفسر	التباين المفسر التراكمي	مؤشر (2)	مؤشر (3)
10	250	1	**5.36**	**53.6**[1]	53.6	**4.5**	**37.4**	**5.36**	**53.6**[1]	53.6	**4.5**	**37.4**
		2	1.20	12.0	65.6			1.20	12.0	65.6		
		3	1.09	10.9	**76.5**[4]			1.09	10.9	**76.5**[4]		
	500	1	**4.99**	**49.9**[1]	49.9	**4.5**	**60.7**	**4.99**	**49.9**[1]	49.9	**4.5**	**60.7**
		2	1.12	11.2	61.2			1.12	11.2	61.2		
		3	1.06	10.6	**71.7**[4]			1.06	10.6	**71.7**[4]		
	1000	1	**4.97**	**49.7**[1]	49.7	**4.6**	**42.8**	**4.97**	**49.7**[1]	49.7	**4.6**	**42.8**
		2	1.07	10.7	60.4			1.07	10.7	60.4		
		3	0.98	9.8	**70.2**[4]			0.98	9.8	**70.2**[4]		
	1500	1	**4.90**	**49.0**[1]	49.0	**4.7**	**53.8**	**4.97**	**49.7**[1]	49.7	**4.6**	**42.8**
		2	1.05	10.5	59.5			1.07	10.7	60.4		
		3	0.98	9.8	**69.3**[4]			0.98	9.8	**70.2**[4]		
40	250	1	**22.46**	**56.1**[1]	56.1	**13.1**	**131.1**	**22.46**	**56.1**[1]	56.1	**13.1**	**131.1**
		2	1.72	4.3	60.4			1.72	4.3	60.4		
		3	1.56	3.9	**64.3**[4]			1.56	3.9	**64.3**[4]		
	500	1	**20.38**	**51.0**[1]	51.0	**14.1**	**245.7**	**20.38**	**51.0**[1]	51.0	**14.1**	**245.7**
		2	1.44	3.6	54.6			1.44	3.6	54.6		
		3	1.36	3.4	**58.0**[4]			1.36	3.4	**58.0**[4]		
	1000	1	**19.35**	**48.4**[1]	48.4	**15.2**	**867.6**	**19.35**	**48.4**[1]	48.4	**15.2**	**867.6**
		2	1.27	3.2	51.6			1.27	3.2	51.6		
		3	1.25	3.1	**54.7**[4]			1.25	3.1	**54.7**[4]		
	1500	1	**18.79**	**47.0**[1]	47.0	**15.3**	**958.2**	**18.79**	**47.0**[1]	47.0	**15.3**	**958.2**
		2	1.23	3.1	50.1			1.23	3.1	50.1		
		3	1.21	3.0	**53.1**[4]			1.21	3.0	**53.1**[4]		

(1): التباين المفسر للمكون الأول يزيد على 20.0%

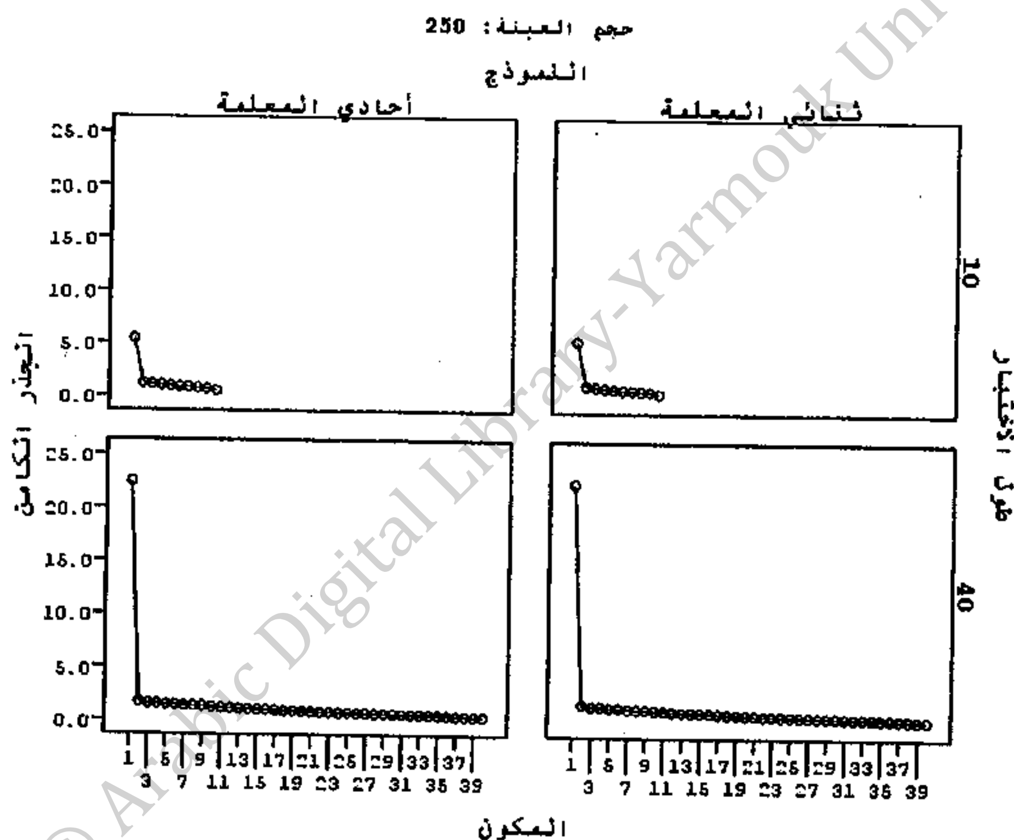
(2): $\frac{\text{الجذر الكامن الأول}}{\text{الجذر الكامن الثاني}}$ يزيد على 2 صحيح

(3): $\frac{(\text{الجذر الكامن الأول} - \text{الجذر الكامن الثاني})}{(\text{الجذر الكامن الثاني} - \text{الجذر الكامن الثالث})}$ عدد كبير يزيد على 7

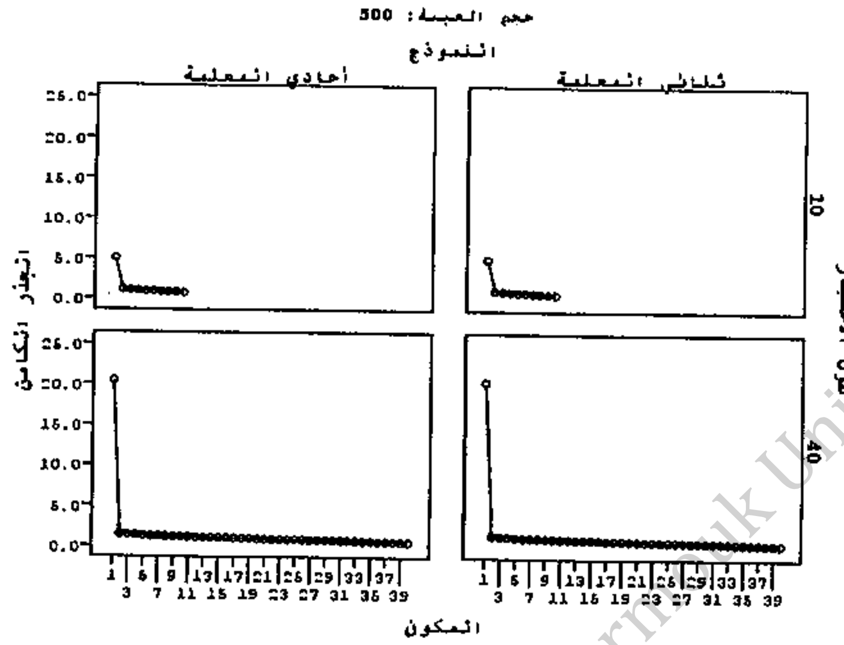
(4): التباين المفسر التراكمي يزيد على 54.0%

يظهر الجدول (2) أن كافة قيم التباين المفسر (المؤشر الأول) عند جميع مستويات متغير عدد الفقرات ومستويات متغير حجم العينة (1500,1000,500,250) قد تخطت قيمة الـ 20% كمؤشر أول لأحادية البعد، وتحقق المؤشر الثاني لأحادية البعد الناتج عن حاصل قسمة الجذر الكامن الأول على الجذر الكامن الثاني حيث تخطت القيم العدد 2، وكذلك أظهرت عملية قسمة حاصل طرح الجذر الكامن الثاني من الجذر الكامن الأول على حاصل طرح الجذر الكامن الثالث من الجذر الكامن الثاني نتائج ضخمة مما يشير إلى تحقق افتراض أحادية البعد كمؤشر

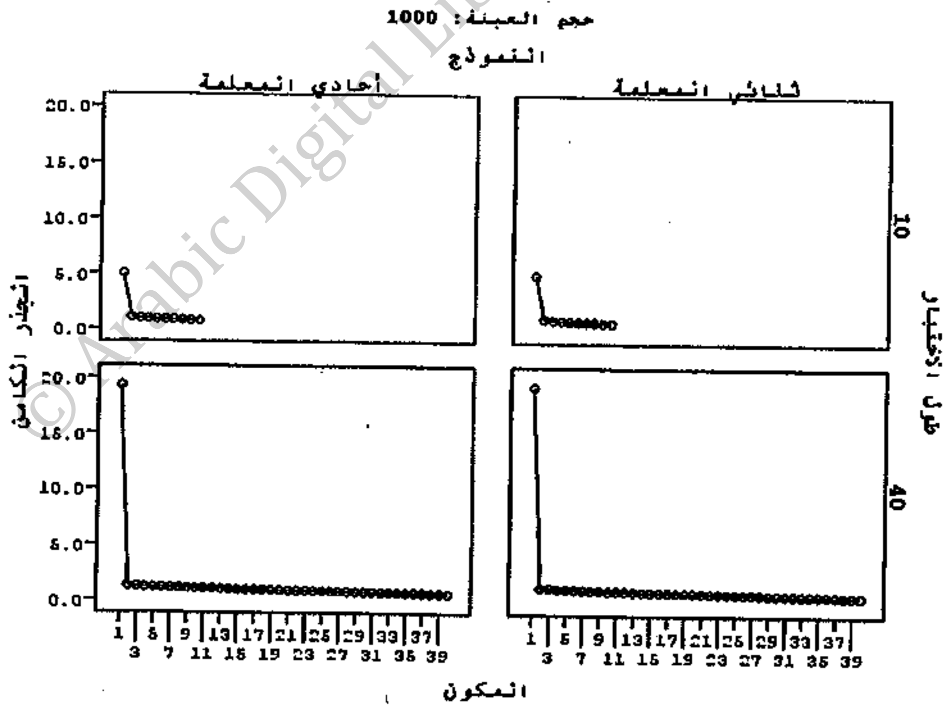
ثالث، ويلاحظ أن قيم التباين المفسر التراكمي تزيد على 54%. والأشكال (2، 3، 4، 5)، تبين نتائج التحليل العاملي للبيانات المولدة (الحقيقية) وفقاً لمتغيرات الدراسة (النموذج اللوغاريتمي، وعدد الفقرات، وحجم العينة) باستخدام الجذور الكامنة وعدد العوامل كمؤشر على تحقق أحادية البعد.(Hattie, 1985)



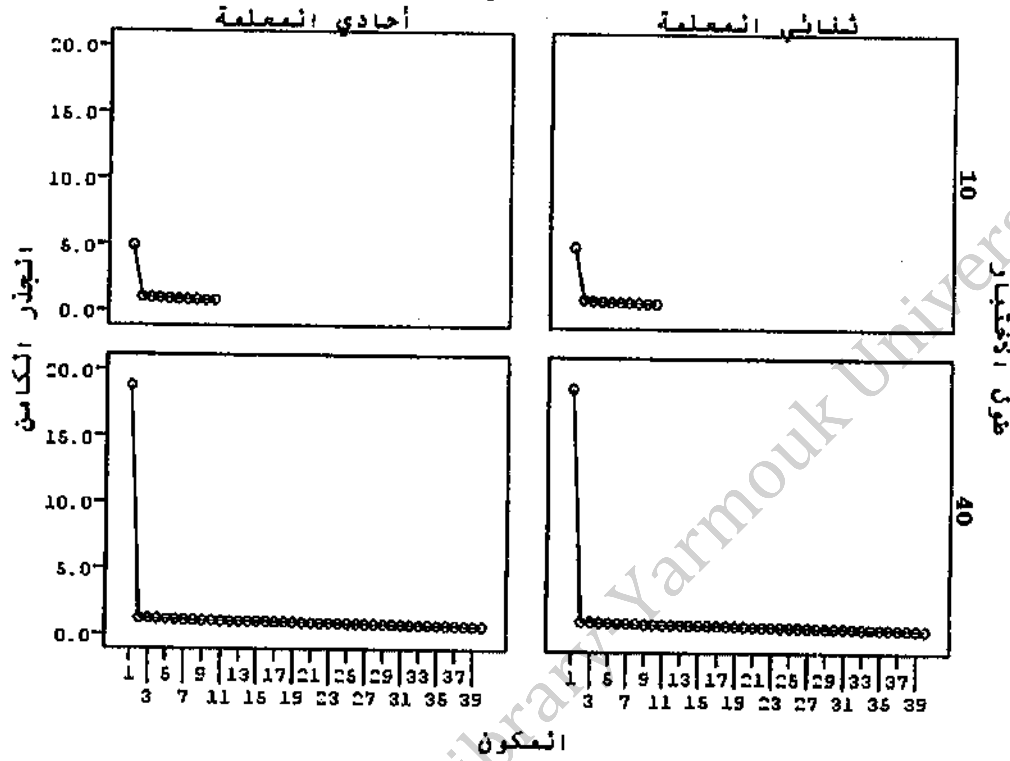
الشكل 2: رسم بياني يوضح نتائج التحليل العاملي للبيانات المولدة وفقا لمتغيرات الدراسة (النموذج اللوغاريتمي، وعدد الفقرات، وعندما يكون حجم العينة 250 فرداً).



الشكل 3: رسم بياني يوضح نتائج التحليل العاملي للبيانات المولدة وفقا لمتغيرات الدراسة (النموذج اللوغاريتمي، وعدد الفقرات، وعندما يكون حجم العينة 500 فرداً).



الشكل 4: رسم بياني يوضح نتائج التحليل العاملي للبيانات المولدة وفقا لمتغيرات الدراسة (النموذج اللوغاريتمي، وعدد الفقرات، وعندما يكون حجم العينة 1000 فرداً).



الشكل 5: رسم بياني يوضح نتائج التحليل العاملي للبيانات المولدة وفقا لمتغيرات الدراسة (النموذج اللوغاريتمي، وعدد الفقرات، وعندما يكون حجم العينة 1500 فرداً).

تقديرات معالم الفقرة للبيانات المولدة:

تم تقدير معالم الفقرات بطريقتين مختلفتين هما: طريقة الأرجحية العظمى الهامشية (MML) باستخدام برنامج (Bilog-MG 3)، وطريقة بييز (Bayes) باستخدام برنامج (WinBUGS v 1. 4).

تقدير معالم الفقرة بطريقة الأرجحية العظمى الهامشية (MML):

برنامج (Bilog-MG3):

استخدم الباحث برنامج Bilog-MG 3 في تقدير معالم الفقرات وفق طريقة (MML)، لما له من مميزات منها : يقدم تقديرات ذات كفاءة عالية مع الفقرة ثنائية الاستجابة، ويفيد في معادلة المجموعات المتكافئة وغير المتكافئة، وفي المعادلة الرأسية، وتحيز الفقرات، وتوزيعات القدرة، وإحصائيات مطابقة الفقرة، والثبات النظري والتجريبي (Zimowski *et. al.* ,2003)، والبرنامج

مزود بمربع حوار لمساعدة الباحث في كتابة هذه الأوامر، بالإضافة إلى وجود مجلد ضمن هذا البرنامج يتضمن العديد من الأمثلة التوضيحية وملفات الأوامر التي قد تفيد مستخدم البرنامج، ويحتوي برنامج Bilog-MG3 على ثلاثة خيارات لطرق تقدير معلمة القدرة للفرد وهي: الأرجحية العظمى الهامشية (Marginal Maximum Likelihood (MML، والبييزي البعدي المتوقع (Expecting a posteriori (EAP والبييزي البعدي العظمى (Maximum a posteriori MAP). أما تقدير معالم الفقرات في برنامج Bilog-MG3 فيتم من خلال طريقة Marginal Maximum Likelihood (MML) ولا يمكن تقديرها بطرق (EAP, MAP) وهذا ما تمت الإشارة إليه في دليل التشغيل لبرنامج Bilog-MG3 (Du Toit, M, 2003).

تقدير معالم الفقرة باستخدام برنامج BILOG-MG3:

استدعيت البيانات التي تم توليدها باستخدام برنامج (WinGen v. 3) لنموذجي نظرية استجابة الفقرة المستخدمة في الدراسة ولكل نموذج بيانات على حدة، أولا النموذج اللوغاريتمي أحادي المعلمة (1-PL)، ثم النموذج اللوغاريتمي ثنائي المعلمة (2-PL)، عند كل مستوى من مستويات متغير حجم العينة، (1500,1000,500,250) وعند كل مستوى من مستويات متغير عدد الفقرات (40,10) وبعد ذلك تم تمريرها إلى برنامج (Bilog-MG3)، من أجل تقدير معالم الفقرات بطريقة (MML) وفق تصميم الأوامر (syntax) للبرنامج وبما يتناسب مع كل حالة من الحالات التي اعتمدت في الدراسة والتي يتم فيها التقدير والمبين في الملحق رقم (1)، و(3).

تقدير معالم الفقرة بطريقة بييز (Bayes):

برنامج (WinBUGS v 1. 4)

إن البرمجيات التي تستخدم منحى الاستدلال البييزي في تقدير معالم الفقرات قليلة جداً، مقارنة مع البرمجيات التي تستخدم طريق الأرجحية العظمى، ويعود سبب ذلك إلى الإجراءات المعقدة لتقديرات بييز الاحتمالية، وهذا ما تم ملاحظته أثناء تقدير معالم الفقرات بطريقة بييز

باستخدام برنامج (WinBUGS v 1. 4) التي تعمل وفق منحى بييز، حيث استغرق تقدير معالم الفقرات لعينة (1500) فرداً ما يقارب (24) ساعة، وعند زيادة حجم العينة أو عدد الفقرات يتطلب وقت أطول للحصول على النتائج، وقد تفشل المحاولة مما يتطلب أجهزة حاسوب متقدمة لتتمكن من القيام بهذه المهمة.

برنامج البق (BUGS) يستخدم الاستدلال البييزي باستخدام عينات جيبس (Gibbs Sampling)، بدأ البرنامج كمشروع للبحوث الإحصائية في مجلس البحوث الطبية في وحدة الإحصاء الحيوي في كامبريدج، المملكة المتحدة عام 1989م، واستمرت عملية التحسين والتطوير للبرنامج حتى تم إصدار الجيل (WinBUGS) الذي يعمل وفق نظام التشغيل (Windows) في عام 1997م، ويتوفر الآن الجيل (WinBUGS v 1.4) الأكثر تحديثا وتطورًا مما سبقه (Cowls, 2004) ويبين الملحق رقم (3) واجهة البرنامج.

تقدير معالم الفقرة باستخدام برنامج WinBUGSv1. 4:

استدعيت البيانات التي تم توليدها باستخدام برنامج (WinGen v. 3) لنموذجي نظرية استجابة الفقرة ولكل نموذج بيانات على حدةٍ، أولاً النموذج اللوغاريتمي أحادي المعلمة (1-PL)، ثم النموذج اللوغاريتمي ثنائي المعلمة (2-PL)، عند كل مستوى من مستويات متغير حجم العينة، (1500,1000,500,250) وعند كل مستوى من مستويات متغير عدد فقرات الاختبار (10 ,40) تم تمريرها لبرنامج (WinBUGS v1. 4) بهدف تقدير معالم الفقرات بطريقة (Bayes) وفق كل موقف من مواقف الدراسة سابقة الذكر.

وتمت عملية تقدير معالم الفقرات بطريقة Bayes وفق الخطوات الآتية:

1. تأمين ملف أوامر syntax يلبي متطلبات النموذج اللوغاريتمي قيد الاهتمام، (انظر الملحق (4)).

2. تأمين ملف بيانات يتناسب مع طبيعة الأوامر المستلمة في الخطوة 1. [انظر الملحق (5)].

3. يتم تظليل الكلمة المفتاحية (Model).

4. يتم فتح قائمة (Model).

5. يتم اختيار أمر (Specifications).

6. يتم إجراء نقرة على زر "فحص النموذج" (check Model).

7. يتم وضع رسالة نصية في سطر الحالة Statusbar تفيد أن النموذج صحيح من الناحية البنائية (Model is Syntadically Corret).

8. يتم الانتقال إلى نافذة ملف البيانات عن طريق قائمة النوافذ (Window).

9. يتم تظليل الكلمة المفتاحية (List) في ملف البيانات.

10. يتم إجراء نقرة على زر "تحميل البيانات" (Load Data).

11. يتم وضع رسالة نصية في سطر الحالة تفيد أن البيانات قد حُمّلت (data loaded).

12. يتم إجراء نقرة على زر "التوليف" (Compile).

13. يتم وضع رسالة نصية في سطر الحالة تفيد أن النموذج قد تم توليفه (Model compiled).

14. يتم إجراء نقرة على زر "تخليق القيم الابتدائية للتوزيعات المسبقة (Gen inits for prior Dis).

15. يتم وضع رسالة نصية في سطر الحالة تفيد أنه "تم تخليق القيم الابتدائية للتوزيعات المسبقة وأنه قد تم تجهيز النموذج، (Initial Values generated Model Initialized).

16. يتم فتح قائمة Ineferance.

17. يتم اختيار الأمر Samples.

18. يتم إدخال الحروف الخاصة بالمعالم حسبما ذكرت في ملف الأوامر مع مراعاة طبيعة النموذج (1-PL, 2-PL) بمعنى ندخل الحرف (a) في إشارة لمعلمة التمييز متبوعة بإجراء نقرة على زر التهيئة (set) فتدخل الحرف (b) في إشارة لمعلمة الصعوبة متبوعة بإجراء نقرة على زر التهيئة (Set) ثم ندخل رمز النجمة (*) في الحقل النصي للعقدة (node) (متبوعة بإجراء نقرة على زر التهيئة (set).

19. يتم فتح قائمة النموذج Model مرة أخرى.

20. يتم اختيار الأمر "التحديث" (Update).

21. يظهر صندوق حواري (Intoactive Dialog Box) عنوانه "اداة التحديث" (Update Tool).

22. يتم وضع قيمة لعدد مرات التحديث في المرحلة التكيفية (Adaptive Stage).

23. يتم النقر على زر "التحديث" (Update).

24. يتم وضع رسالة نصية في سطر الحالة تفيد أن المرحلة التكيفية قد نفذت في كذا ثانية.

25. يتم وضع قيمة لعدد مرات التحديث (ويفضل أن تزيد على القيمة المذكورة في النقطة (22)) لمرحلة التسخين (Burn In stage).

26. يتم النقر على زر التحديث.

27.يتم وضع رسالة نصية في سطر الحالة تفيد أن مرحلة التسخين قد نفذت في كذا ثانية.

28.يتم وضع قيمة لعدد مرات التحديث (يفضل أن تكون من مضاعفات القيمة الموضوعة في النقطة (25)) لمرحلة تقدير المعالم بالشكل النهائي. (Final estimating Stage).

29.يتم وضع رسالة نصية في سطر الحالة تفيد أن مرحلة تقدير المعالم النهائية قد نفذت في كذا ثانية وهي مرحلة تستغرق الكثير من الزمن حيث وصلت إلى 26 ساعة في حالة النموذج ثنائي المعلمة (40) فقرة و(1500) فرداً.

30.يتم النقر على زر الإحصائيات (stats) بهدف عرض نتائج عملية التقدير لمعالم الفقرات.

31.يتم فتح قائمة الملف File وتخزين الإحصائيات في ملف خاص بها ليتم التعامل معه في اتمام بقية الإجراءات اللازمة للدراسة.

المعالجات الإحصائية:

للإجابة عن تساؤلات الدراسة تم توليد بيانات الدراسة باستخدام برنامج (WinGen-3) وفقاً لمتغيرات الدراسة (النموذج اللوغاريتمي، وحجم العينة، وعدد الفقرات) ثم تم استدعاء هذه البيانات لإجراء التحليلات باســتخدام برمجيــات (Bilog-MG3) و(WinBUGS V 1.4) و(SPSS).

وللإجابة عن أسئلة الدراسة تم إجراء المعالجات الإحصائية الآتية:

1) الوسط الحسابي والانحراف المعياري الخاصة بمعالم الفقــرات باســتخدام طريقتــي التقــدير (Bayes, MML).

2) مؤشر دقة التقدير (RMSE) حيث تم حساب الأوساط الحسابية للمؤشر والخاصة بمعالم الفقرات باستخدام كل طريقة من طرق التقدير (Bayes, MML) .

$$RMSE_x = \sqrt{\frac{\sum_{i=1}^{n} (\hat{X}_{estimated} - X_{true})^2}{n}}$$

3) حساب التباين للقيم المقدرة لمعلمة الصعوبة وفقاً لطريقتي التقدير (Bayes, MML) في المواقف المختلفة للمتغيرات (النموذج اللوغاريتمي المستخدم وعدد الفقرات وحجم العينة) ثم تمت قسمة تباين القيم المقدرة لمعلمة الصعوبة وفقاً لطريقة التقدير Bayes على تباين القيم المقدرة لمعلمة الصعوبة وفقاً لطريقة التقدير MML كمؤشر كفاءة نسبية في تقدير معلمة الصعوبة للفقرة في المواقف المختلفة وفقاً لطريقتي التقدير والتي تأخذ الصيغة الرياضية $RE_b = \frac{V\ \hat{b}_{Bayes}}{V\ \hat{b}_{MML}}$ حيث أن الرمز V يُعني تباين تقديرات معلمة الصعوبة وفقاً لطريقة التقدير المعنية.

4) حساب التباين للقيم المقدرة لمعلمة التمييز وفقاً لطريقتي التقدير (Bayes, MML) في المواقف المختلفة للمتغيرات (النموذج اللوغاريتمي المستخدم وعدد الفقرات وحجم العينة) ثم تمت قسمة تباين القيم المقدرة لمعلمة التمييز وفقاً لطريقة التقدير Bayes على تباين القيم المقدرة لمعلمة التمييز وفقاً لطريقة التقدير MML كمؤشر كفاءة نسبية في تقدير معلمة التمييز للفقرة في المواقف المختلفة وفقاً لطريقتي التقدير والتي تأخذ الصيغة الرياضية $RE_a = \frac{V\ \hat{a}_{Bayes}}{V\ \hat{a}_{MML}}$ ، حيث أن الرمز V يُعني تباين تقديرات معلمة التمييز وفقاً لطريقة التقدير المعنية.

5) معامل ارتباط بيرسون: استخدم معامل الارتباط لتحديد قوة العلاقة بين القيم المقدرة لكل طريقة من طرق التقدير مع القيم المولدة (الحقيقية)، كمؤشر من مؤشرات دقة التقدير لمعالم الفقرات.

6) اختبار (t-test) لمعاملات الارتباط للعينات غير المستقلة: بهدف معرفة الدلالة الإحصائية لمعاملات الارتباط بين القيم المقدرة والقيم الحقيقية وفق كل طريقة من طريقتي التقدير.

الفصل الرابع

عرض نتائج الدراسة

يتناول هذا الفصل عرضاً لنتائج الدراسة مبوبة حسب أسئلتها:

تقدير معالم الفقرات باستخدام طريقة التقدير (MML, Bayes) باختلاف المتغيرات (النموذج، عدد الفقرات، حجم العينة).

تم حساب الأوساط الحسابية والانحرافات المعيارية لتقديرات معالم الفقرة وأخطاءها المعيارية المقدرة وفقاً لطريقتي التقدير (MML, Bayes) باختلاف المتغيرات (النموذج، عدد الفقرات، حجم العينة)، وذلك كما هو مبين في الجدول (3).

جدول (3)

الأوساط الحسابية لتقديرات معالم الفقرة وأخطاؤها المعيارية المقدرة وفقاً للطريقتين (MML, Bayes) باختلاف المتغيرات (النموذج، عدد الفقرات، حجم العينة)

النموذج	حجم العينة	الإحصائي	عدد الفقرات: 10				عدد الفقرات: 40			
			MML		Bayes		MML		Bayes	
			الوسط الحسابي	الانحراف المعياري	الوسط الحسابي	الانحراف المعياري	الوسط الحسابي	الانحراف المعياري	الوسط الحسابي	الانحراف المعياري
أحادي المعلمة	250	معلمة الصعوبة	0.000	0.63	-0.473	0.64	0.000	0.82	-0.536	0.80
		الخطأ المعياري للصعوبة	0.20	0.01	0.15	0.01	0.20	0.03	0.16	0.02
	500	معلمة الصعوبة	0.000	0.64	-0.416	0.66	0.000	0.80	-0.610	0.79
		الخطأ المعياري للصعوبة	0.15	0.01	0.11	0.01	0.15	0.02	0.11	0.01
	1000	معلمة الصعوبة	0.000	0.70	-0.557	0.73	0.000	0.77	-0.707	0.78
		الخطأ المعياري للصعوبة	0.10	0.01	0.08	0.01	0.10	0.01	0.08	0.01
	1500	معلمة الصعوبة	0.000	0.69	-0.502	0.72	0.000	0.80	-0.730	0.81
		الخطأ المعياري للصعوبة	0.08	0.01	0.06	0.00	0.09	0.01	0.07	0.01
ثنائي المعلمة	250	معلمة التمييز	0.870	0.45	1.024	0.62	0.800	0.24	1.107	0.37
		الخطأ المعياري للتمييز	0.16	0.05	0.27	0.18	0.16	0.04	0.23	0.06
		معلمة الصعوبة	-0.467	0.89	-0.398	0.72	-0.883	1.08	-0.456	0.69
		الخطأ المعياري للصعوبة	0.24	0.12	0.23	0.07	0.29	0.18	0.20	0.05
	500	معلمة التمييز	0.740	0.25	0.801	0.28	0.811	0.26	0.962	0.33
		الخطأ المعياري للتمييز	0.11	0.02	0.17	0.04	0.12	0.04	0.15	0.04
		معلمة الصعوبة	-0.425	0.92	-0.390	0.79	-0.779	0.92	-0.657	0.74
		الخطأ المعياري للصعوبة	0.19	0.09	0.19	0.05	0.19	0.08	0.17	0.04
	1000	معلمة التمييز	0.781	0.30	0.843	0.32	0.817	0.26	0.892	0.28
		الخطأ المعياري للتمييز	0.08	0.02	0.12	0.03	0.09	0.02	0.10	0.02
		معلمة الصعوبة	-0.483	1.20	-0.497	0.97	-0.888	0.91	-0.729	0.80
		الخطأ المعياري للصعوبة	0.16	0.15	0.15	0.07	0.15	0.07	0.13	0.05
	1500	معلمة التمييز	0.740	0.24	0.784	0.25	0.769	0.26	0.842	0.28
		الخطأ المعياري للتمييز	0.06	0.01	0.10	0.02	0.07	0.02	0.08	0.02
		معلمة الصعوبة	-0.423	1.16	-0.439	0.99	-0.888	0.90	-0.764	0.83
		الخطأ المعياري للصعوبة	0.13	0.11	0.13	0.08	0.12	0.04	0.12	0.03

يلاحظ من الجدول (3)، أن النتائج الخاصة به قد كانت على النحو الآتي:

أ. **فيما يخص الوسط الحسابي:** إن كافة قيم الوسط الحسابي لتقديرات معلمة الصعوبة على اختلاف عدد الفقرات (10 أو 40) وعلى اختلاف حجم العينة (250، 500، 1000، 1500)

في النموذج الأحادي (1-pl) حسب طريقة التقدير MML قد كانت **أصغر** ظاهرياً من نظيراتها حسب طريقة التقدير Bayes، كما يلاحظ من الجدول (3)، أن كافة قيم الوسط الحسابي لتقديرات معلمة التمييز على اختلاف عدد الفقرات (10 أو 40) وعلى اختلاف حجم العينة (250، 500، 1000، 1500) في النموذج الثنائي (2-pl) حسب طريقة التقدير MML قد كانت **أصغر** ظاهرياً من نظيراتها حسب طريقة التقدير Bayes، في حين أن قيم الوسط الحسابي لتقديرات معلمة الصعوبة عندما يكون عدد الفقرات 40 فقرة وعلى اختلاف حجم العينة (250، 500، 1000، 1500) في النموذج الثنائي (2-pl) حسب طريقة التقدير MML قد كانت **أكبر** ظاهرياً من نظيراتها حسب طريقة التقدير Bayes، كما أن قيم الوسط الحسابي لتقديرات معلمة الصعوبة عندما يكون عدد الفقرات 10 فقرات وعندما يكون حجم العينة (250، 500) في النموذج الثنائي (2-pl) حسب طريقة التقدير MML قد كانت **أكبر** ظاهرياً من نظيراتها حسب طريقة التقدير Bayes، وأخيراً أن قيم الوسط الحسابي لتقديرات معلمة الصعوبة عندما يكون عدد الفقرات 10 فقرات وعندما يكون حجم العينة (1000، 1500) في النموذج الثنائي (2-pl) حسب طريقة التقدير MML قد كانت **أصغر** ظاهرياً من نظيراتها حسب طريقة التقدير Bayes.

ب. **فيما يخص الوسط الحسابي للخطأ المعياري:** يلاحظ من الجدول (5)، أن كافة قيم الأوساط الحسابية للأخطاء المعيارية لتقديرات معلمة الصعوبة على اختلاف عدد الفقرات (10 أو 40) وعلى اختلاف حجم العينة (250، 500، 1000، 1500) في النموذج الأحادي (1-pl) حسب طريقة التقدير MML قد كانت **أكبر** ظاهرياً من نظيراتها حسب طريقة التقدير Bayes، كما يلاحظ من الجدول (5)، أن كافة قيم الأوساط الحسابية للأخطاء المعيارية لتقديرات معلمة التمييز على اختلاف عدد الفقرات (10 أو 40) وعلى اختلاف حجم العينة (250، 500،

1000، 1500) في النموذج الثنائي (2-pl) حسب طريقة التقدير MML قد كانت **أصغر** ظاهرياً من نظيراتها حسب طريقة التقدير Bayes، في حين أن كافة قيم الأوساط الحسابية للأخطاء المعيارية لتقديرات معلمة الصعوبة باختلاف عدد الفقرات (10، 40) فقرة باختلاف حجم العينة (250، 500، 1000، 1500) في النموذج الثنائي (2-pl) حسب طريقة التقدير MML قد كانت **أكبر** ظاهرياً من نظيراتها حسب طريقة التقدير Bayes، باستثناء أن قيم الوسط الحسابي للخطأ المعياري لتقديرات معلمة الصعوبة عندما يكون عدد الفقرات 10 فقرات وحجم العينة500 فرداً في النموذج الثنائي (2-pl) حسب طريقة التقدير MML قد كانت **أصغر** ظاهرياً من نظيراتها حسب طريقة التقدير Bayes.

أولاً. النتائج المتعلقة بالإجابة عن سؤال الدراسة الأول:

"ما أثر طريقة التقدير (MML، وBayes) في دقة تقدير معالم الفقرة في ضوء اختلاف المتغيرات (النموذج اللوغاريتمي المستخدم وعدد الفقرات، وحجم العينة)؟"

للإجابة عن هذا السؤال؛ فقد تم حساب الأوساط الحسابية والانحرافات المعيارية الخاصة بمؤشر RMSE كمؤشر من مؤشرات دقة التقدير لمعلمة التمييز للنموذج الثنائي (2-pl) باستخدام المعادلة $RMSE_a = \sqrt{\frac{\sum_{i=1}^{n} (\hat{a}_{estimated} - a_{true})^2}{n}}$ باختلاف طريقة التقدير (MML، Bayes)، وفقاً للمتغيرين (عدد الفقرات، حجم العينة)، وذلك كما هو مبين في الجدول (4).

جدول 4

الأوساط الحسابية لمؤشر RMSE في دقة تقدير معلمة التمييز للنموذج الثنائي باختلاف طريقة التقدير (MML، Bayes)، واختلاف المتغيرين (عدد الفقرات، حجم العينة)

طول الاختبار	مؤشر دقة التقدير RMSE لمعلمة التمييز وفقاً لطريقة التقدير:	حجم العينة 250	500	1000	1500
10	BAYES	0. 478	0. 344	0. 317	0. 363
	MML	0. 399	0. 394	0. 369	0. 403
40	BAYES	0. 315	0. 315	0. 262	0. 317
	MML	0. 319	0. 335	0. 297	0. 343

يلاحظ من الجدول (4)، أن كافة قيم مؤشر RMSE كمؤشر دقة لتقديرات معلمة التمييز للنموذج الثنائي حسب طريقة التقدير MML باختلاف عدد الفقرات (10، 40) واختلاف حجم العينة

(250، 500، 1000، 1500) قد كانت أكبر ظاهرياً من نظيراتها حسب طريقة التقــدير Bayes باستثناء أن قيمة مؤشر RMSE كمؤشر دقة لتقديرات معلمة التمييز للنموذج الثنائي حسب طريقة التقدير MML عندما كان عدد الفقرات 10 فقرات وحجم عينة 250 قد كانت **أصــغر** ظاهرياً مــن نظيرتها حسب طريقة التقدير بييز (Bayes).

وكذلك تم حساب الأوساط الحسابية والانحرافات المعيارية الخاصة بمؤشر RMSE كمؤشر من مؤشرات دقة التقدير لمعلمة الصعوبة باستخدام المعادلــة $RMSE_b=\sqrt{\frac{\sum_{i=1}^{n}(\hat{b}_{estimated}-b_{true})^2}{n}}$ باختلاف طريقة التقدير (MML، وBayes)، واختلاف المتغيرات (النموذج، عدد الفقرات، حجم العينة)، وذلك كما هو مبين في الجدول (5).

جدول (5)

الأوساط الحسابية لمؤشر RMSE في دقة تقدير معلمة الصعوبة باختلاف طريقة التقدير (MML، وBayes)، وباختلاف المتغيرات (النموذج، عدد الفقرات، حجم العينة)

النموذج	طول الاختبار	مؤشر دقة التقدير RMSE لمعلمة الصعوبة وفقاً لطريقة التقدير:	حجم العينة 250	500	1000	1500
أحادي المعلمة	10	BAYES	0. 803	0. 785	0. 865	0. 803
		MML	0. 475	0. 530	0. 460	0. 442
	40	BAYES	0. 507	0. 549	0. 646	0. 647
		MML	0. 365	0. 352	0. 384	0. 352
ثنائي المعلمة	10	BAYES	0. 741	0. 692	0. 767	0. 676
		MML	0. 764	0. 696	0. 786	0. 680
	40	BAYES	0. 539	0. 560	0. 672	0. 691
		MML	0. 812	0. 689	0. 787	0. 774

يلاحظ من الجدول (5)، أن كافة قيم مؤشر RMSE كمؤشر دقة لتقديرات معلمة الصــعوبة للنموذج الأحادي حسب طريقة التقدير MML باختلاف عدد الفقرات (10، 40) واخــتلاف حجــم العينة (250، 500، 1000، 1500) قد كانت **أصغر** ظاهرياً من نظيراتها حسب طريقة التقــدير Bayes؛ مما يُعني أن طريقة MML كانت أكثر دقة ظاهرياً في تقدير معلمة الصعوبة من نظيرتها طريقة Bayes في تقدير معلمة الصعوبة.

كما يلاحظ من الجدول (5)، أن كافة قيم مؤشر RMSE كمؤشر دقة لتقديرات معلمة الصعوبة للنموذج الثنائي حسب طريقة التقدير MML باختلاف عدد الفقرات (10، 40) واختلاف حجم العينة (250، 500، 1000، 1500) قد كانت أكبر ظاهرياً من نظيراتها حسب طريقة التقدير Bayes.

النتائج المتعلقة بالإجابة عن سؤال الدراسة الثاني:

"ما أثر طريقة التقدير (MML، وBayes) في الكفاءة النسبية لتقدير معالم الفقرة في ضوء اختلاف المتغيرات (النموذج اللوغاريتمي المستخدم وعدد الفقرات، وحجم العينة)؟"

للإجابة عن سؤال الدراسة الثاني، فقد تم حساب التباين للقيم المقدرة لمعلمة الصعوبة وفقاً لطريقتي التقدير (Bayes, MML) في المواقف المختلفة للمتغيرات (النموذج اللوغاريتمي المستخدم وعدد الفقرات وحجم العينة) ثم تمت قسمة تباين القيم المقدرة لمعلمة الصعوبة وفقاً لطريقة التقدير Bayes على تباين القيم المقدرة لمعلمة الصعوبة وفقاً لطريقة التقدير MML كمؤشر كفاءة نسبية في تقدير معلمة الصعوبة للفقرة في المواقف المختلفة وفقاً لطريقتي التقدير والتي تأخذ الصيغة الرياضية $RE_b = \frac{V\hat{b}_{Bayes}}{V\hat{b}_{MML}}$، وذلك كما هو مبين في الجدول (6).

الجدول (6)

قيم الكفاءة النسبية لمعلمة الصعوبة وفقاً لطريقتي التقدير (Bayes, MML) في المواقف المختلفة للمتغيرات (النموذج اللوغاريتمي المستخدم وعدد الفقرات وحجم العينة)

النموذج	طول الاختبار	الإحصائي	حجم العينة 250	500	1000	1500
أحادي المعلمة	10	تباين معلمة الصعوبة وفق طريقة BAYES	0.411	0.438	0.532	0.525
		تباين معلمة الصعوبة وفق طريقة MML	0.397	0.404	0.493	0.477
		الكفاءة النسبية لـ BAYES على MML	**1.036**	**1.083**	**1.079**	**1.101**
	40	تباين معلمة الصعوبة وفق طريقة BAYES	0.644	0.628	0.603	0.652
		تباين معلمة الصعوبة وفق طريقة MML	0.680	0.637	0.596	0.638
		الكفاءة النسبية لـ BAYES على MML	**0.947**	**0.986**	**1.012**	**1.023**
ثنائي المعلمة	10	تباين معلمة الصعوبة وفق طريقة BAYES	0.519	0.623	0.940	0.976
		تباين معلمة الصعوبة وفق طريقة MML	0.786	0.845	1.437	1.350
		الكفاءة النسبية لـ BAYES على MML	**0.660**	**0.738**	**0.654**	**0.723**
	40	تباين معلمة الصعوبة وفق طريقة BAYES	0.480	0.550	0.640	0.688
		تباين معلمة الصعوبة وفق طريقة MML	1.160	0.842	0.830	0.811
		الكفاءة النسبية لـ BAYES على MML	**0.414**	**0.653**	**0.770**	**0.849**

يلاحظ من الجدول (6)، أن قيم الكفاءة النسبية لطريقة Bayes على MML قد كانت ظاهرياً لصالح طريقة التقدير MML عندما يكون عدد الفقرات 10 على اختلاف حجوم العينات في النموذج اللوغاريتمي الأحادي، وأن قيم الكفاءة النسبية لطريقة Bayes على MML قد كانت ظاهرياً لصالح طريقة التقدير MML عندما يكون عدد الفقرات 40 في حجوم العينات (1000، 1500) في النموذج اللوغاريتمي الأحادي، وعلى العكس من ذلك فإن قيم الكفاءة النسبية لطريقة Bayes على MML قد كانت ظاهرياً لصالح طريقة التقدير Bayes عندما يكون عدد الفقرات 40 في حجوم العينات (250، 500) في النموذج اللوغاريتمي الأحادي. في حين يلاحظ من الجدول (6)، أن قيم الكفاءة النسبية لطريقة Bayes على MML قد كانت ظاهرياً لصالح طريقة التقدير Bayes عندما يكون عدد الفقرات (10، 40) على اختلاف حجوم العينات في النموذج اللوغاريتمي الثنائي.

وكذلك للإجابة عن سؤال الدراسة الثاني، فقد تم حساب التباين للقيم المقدرة لمعلمة التمييز وفقاً لطريقتي التقدير (Bayes, MML) في المواقف المختلفة للمتغيرات (النموذج اللوغاريتمي المستخدم وعدد الفقرات وحجم العينة) ثم تمت قسمة تباين القيم المقدرة لمعلمة التمييز وفقاً لطريقة التقدير Bayes على تباين القيم المقدرة لمعلمة التمييز وفقاً لطريقة التقدير MML كمؤشر كفاءة نسبية في تقدير معلمة التمييز للفقرة في المواقف المختلفة وفقاً لطريقتي التقدير والتي تأخذ الصيغة الرياضية $RE_a = \frac{V\hat{a}_{Bayes}}{V\hat{a}_{MML}}$، وذلك كما هو مبين في الجدول (7).

الجدول (7)

قيم الكفاءة النسبية لمعلمة التمييز وفقاً لطريقتي التقدير (Bayes, MML) في المواقف المختلفة للمتغيرات (النموذج اللوغاريتمي المستخدم وعدد الفقرات وحجم العينة)

طول الاختبار	الإحصائي	حجم العينة 250	500	1000	1500
10	تباين معلمة التمييز وفق طريقة BAYES	0.384	0.079	0.104	0.062
	تباين معلمة التمييز وفق طريقة MML	0.204	0.061	0.091	0.056
	الكفاءة النسبية لـ BAYES على MML	**1.881**	**1.307**	**1.146**	**1.100**
40	تباين معلمة التمييز وفق طريقة BAYES	0.137	0.110	0.081	0.077
	تباين معلمة التمييز وفق طريقة MML	0.058	0.067	0.070	0.069
	الكفاءة النسبية لـ BAYES على MML	**2.380**	**1.652**	**1.166**	**1.115**

يلاحظ من الجدول (7)، أن قيم الكفاءة النسبية لطريقة Bayes على MML قد كانت ظاهرياً لصالح طريقة التقدير MML عندما يكون عدد الفقرات (10، 40) على اختلاف حجوم العينات في النموذج اللوغاريتمي الثنائي.

النتائج المتعلقة بالإجابة عن سؤال الدراسة الثالث:

"هل توجد فروق دالة إحصائيا عند مستوى الدلالة (α=0. 05) في قيم معاملات الارتباط بين معالم الفقرة الحقيقية وبين معالم الفقرة المقدرة وفق طريقتي التقدير (MML, Bayes) باختلاف المتغيرات (النموذج اللوغاريتمي المستخدم، عدد الفقرات، حجم العينة)؟"

للإجابة عن سؤال الدراسة الثالث، فقد تم حساب قيم معاملات الارتباط بين معلمة التمييز الحقيقية وبين معلمة التمييز المقدرة للنموذج الثنائي وفق طريقتي التقدير (MML، Bayes) باختلاف المتغيرين (عدد الفقرات، حجم العينة)، ثم تم استخدام اختبار t لمعاملات الارتباط للعينات المترابطة للكشف عن جوهرية الفرق في قيمتي (معامل الارتباط بين معلمة التمييز الحقيقية وبين معلمة التمييز المقدرة للنموذج الثنائي وفق طريقة التقدير MML) من جهة و(معامل الارتباط بين معلمة التمييز الحقيقية وبين معلمة التمييز المقدرة للنموذج الثنائي وفق طريقة التقدير Bayes) من جهة أخرى عند كل موقف تفاعلي للمتغيرين (عدد الفقرات، حجم العينة)، وذلك كما هو مبين في الجدول (8).

الجدول (8)

نتائج اختبار t لمعاملات الارتباط للعينات المترابطة لقيم معاملات الارتباط بين معلمة التمييز الحقيقية وبين معلمة التمييز المقدرة للنموذج الثنائي وفق طريقتي التقدير باختلاف المتغيرين (عدد الفقرات، حجم العينة)

عدد الفقرات	حجم العينة	الارتباط بين معلمة التمييز الحقيقية وبين معلمة التمييز المقدرة وفق طريقة التقدير: MML	BAYES	الارتباط لمعلمة التمييز وفق الطريقتين	قيمة ت المحسوبة	الدلالة الإحصائية
10	250	0.74	0.73	0.99	0.474	0.636
	500	0.83	0.85	0.98	-5.307	0.000
	1000	0.84	0.88	0.99	-25.598	0.000
	1500	0.72	0.75	0.99	-14.851	0.000
40	250	0.60	0.56	0.94	2.549	0.011
	500	0.50	0.46	0.98	6.576	0.000
	1000	0.69	0.66	0.99	7.696	0.000
	1500	0.55	0.54	1.00	9.125	0.000

يتضح من الجدول (8)، وجود فرق دال إحصائياً عند مستوى الدلالة α=0.05 عندما كان عدد الفقرات 10 بين قيمتي (معامل الارتباط بين معلمة التمييز الحقيقية وبين معلمة التمييز المقدرة للنموذج الثنائي وفق طريقة التقدير MML) من جهة و(معامل الارتباط بين معلمة التمييز الحقيقية وبين معلمة التمييز المقدرة للنموذج الثنائي وفق طريقة التقدير Bayes) من جهة أخرى عندما كان عدد الفقرات 40 فقرة وباختلاف حجم العينة (250، 500، 1000، 1500) لصالح (معامل الارتباط بين معلمة التمييز الحقيقية وبين معلمة التمييز المقدرة للنموذج الثنائي وفق طريقة التقدير MML). كما يتضح من الجدول (8) وجود فرق دال إحصائياً عند مستوى الدلالة α=0.05 بين قيمتي (معامل الارتباط بين معلمة التمييز الحقيقية وبين معلمة التمييز المقدرة للنموذج الثنائي وفق طريقة التقدير MML) من جهة و(معامل الارتباط بين معلمة التمييز الحقيقية وبين معلمة التمييز المقدرة للنموذج الثنائي وفق طريقة التقدير Bayes) من جهة أخرى عندما كان عدد الفقرات 10 فقرة وباختلاف حجم العينة (500، 1000، 1500) لصالح (معامل الارتباط بين معلمة التمييز الحقيقية وبين معلمة التمييز المقدرة للنموذج الثنائي وفق طريقة التقدير Bayes). في حين لم يثبت من الجدول (8) وجود فرق دال إحصائياً عند مستوى الدلالة α=0.05 بين قيمتي (معامل الارتباط بين معلمة التمييز الحقيقية وبين معلمة التمييز المقدرة للنموذج الثنائي وفق طريقة التقدير MML) من جهة و(معامل الارتباط بين معلمة التمييز الحقيقية وبين معلمة التمييز المقدرة للنموذج الثنائي وفق طريقة التقدير Bayes) من جهة أخرى عندما كان عدد الفقرات 10 فقرة وحجم العينة 250 فرداً.

كما تم حساب قيم معاملات الارتباط بين معلمة الصعوبة الحقيقية وبين معلمة الصعوبة المقدرة وفق طريقتي التقدير (MML، Bayes) باختلاف المتغيرات (النموذج، عدد الفقرات، حجم العينة) ثم تم استخدام اختبار t لمعاملات الارتباط للعينات المترابطة للكشف عن جوهرية الفرق في

قيمتي (معامل الارتباط بين معلمة الصعوبة الحقيقية وبين معلمة الصعوبة المقدرة وفق طريقة التقدير MML) من جهة و(معامل الارتباط بين معلمة الصعوبة الحقيقية وبين معلمة الصعوبة المقدرة وفق طريقة التقدير Bayes) من جهة أخرى عند كل موقف تفاعلي للمتغيرات (النموذج، عدد الفقرات، حجم العينة)، وذلك كما هو مبين في الجدول (9).

الجدول (9)

نتائج اختبار t لمعاملات الارتباط للعينات المترابطة لقيم معاملات الارتباط بين معلمة الصعوبة الحقيقية وبين معلمة الصعوبة المقدرة وفق طريقتي التقدير باختلاف المتغيرات (النموذج، عدد الفقرات، حجم العينة)

طول الاختبار	النموذج	حجم العينة	الارتباط بين معلمة الصعوبة الحقيقية وبين معلمة الصعوبة المقدرة وفق طريقة التقدير: MML	BAYES	الارتباط لمعلمة الصعوبة وفق الطريقتين	قيمة ت المحسوبة	الدلالة الإحصائية
10	أحادي المعلمة	250	0.9765	0.9772	1.00	-9.043	0.000
		500	0.9267	0.9285	1.00	-29.921	0.000
		1000	0.9527	0.9538	1.00	-48.638	0.000
		1500	0.9702	0.9710	1.00	-56.147	0.000
	ثنائي المعلمة	250	0.9423	0.9403	1.00	1.803	0.073
		500	0.9649	0.9627	1.00	3.027	0.003
		1000	0.9523	0.9579	1.00	-8.701	0.000
		1500	0.9810	0.9863	1.00	-20.529	0.000
40	أحادي المعلمة	250	0.9687	0.9688	1.00	-0.247	0.805
		500	0.9809	0.9818	1.00	-8.642	0.000
		1000	0.9746	0.9748	1.00	-4.380	0.000
		1500	0.9807	0.9810	1.00	-8.558	0.000
	ثنائي المعلمة	250	0.9386	0.9495	0.99	-3.500	0.001
		500	0.9617	0.9625	1.00	-0.941	0.347
		1000	0.9618	0.9623	1.00	-0.924	0.356
		1500	0.9605	0.9642	1.00	-15.723	0.000

يتضح من الجدول (9)، وجود فرق دال إحصائياً عند مستوى الدلالة α=0.05 بين قيمتي (معامل الارتباط بين معلمة الصعوبة الحقيقية وبين معلمة الصعوبة المقدرة للنموذج الأحادي وفق طريقة التقدير MML) من جهة و(معامل الارتباط بين معلمة الصعوبة الحقيقية وبين معلمة الصعوبة المقدرة للنموذج الأحادي وفق طريقة التقدير Bayes) من جهة أخرى عندما كان عدد الفقرات 10 فقرة وباختلاف حجم العينة (250، 500، 1000، 1500) لصالح (معامل الارتباط بين معلمة الصعوبة الحقيقية وبين معلمة الصعوبة المقدرة للنموذج الأحادي وفق طريقة التقدير Bayes). كما يتضح من الجدول(9)، وجود فرق دال إحصائياً عند مستوى الدلالة α=0.05 بين

قيمتي (معامل الارتباط بين معلمة الصعوبة الحقيقية وبين معلمة الصعوبة المقدرة للنموذج الأحادي وفق طريقة التقدير MML) من جهة و(معامل الارتباط بين معلمة الصعوبة الحقيقية وبين معلمة الصعوبة المقدرة للنموذج الأحادي وفق طريقة التقدير Bayes) من جهة أخرى عندما كان عدد الفقرات 40 فقرة وباختلاف حجم العينة (500، 1000، 1500) لصالح (معامل الارتباط بين معلمة الصعوبة الحقيقية وبين معلمة الصعوبة المقدرة للنموذج الأحادي وفق طريقة التقدير Bayes)، في حين لم يثبت وجود فرق دال إحصائياً عند مستوى الدلالة α=0.05 بين قيمتي (معامل الارتباط بين معلمة الصعوبة الحقيقية وبين معلمة الصعوبة المقدرة للنموذج الأحادي وفق طريقة التقدير MML) من جهة و(معامل الارتباط بين معلمة الصعوبة الحقيقية وبين معلمة الصعوبة المقدرة للنموذج الأحادي وفق طريقة التقدير Bayes) من جهة أخرى عندما كان عدد الفقرات 40 فقرة وحجم العينة 250 فرداً.

كما يتضح من الجدول (9)، وجود فرق دال إحصائياً عند مستوى الدلالة α=0.05 بين قيمتي (معامل الارتباط بين معلمة الصعوبة الحقيقية وبين معلمة الصعوبة المقدرة للنموذج الثنائي وفق طريقة التقدير MML) من جهة و(معامل الارتباط بين معلمة الصعوبة الحقيقية وبين معلمة الصعوبة المقدرة للنموذج الثنائي وفق طريقة التقدير Bayes، من جهة أخرى عندما كان عدد الفقرات 40 فقرة وباختلاف حجم العينة (250، 1500) لصالح (معامل الارتباط بين معلمة الصعوبة الحقيقية وبين معلمة الصعوبة المقدرة للنموذج الأحادي وفق طريقة التقدير Bayes). في حين لم يثبت وجود فرق دال إحصائياً عند مستوى الدلالة α=0.05 بين قيمتي (معامل الارتباط بين معلمة الصعوبة الحقيقية وبين معلمة الصعوبة المقدرة للنموذج الثنائي وفق طريقة التقدير MML) من جهة و(معامل الارتباط بين معلمة الصعوبة الحقيقية وبين معلمة الصعوبة المقدرة للنموذج الثنائي وفق طريقة التقدير Bayes) من جهة أخرى عندما كان عدد الفقرات 40 فقرة وحجم العينة

(500، 1000) فردا. كما يتضح من الجدول (9)، وجود فرق دال إحصائياً عند مستوى الدلالة α=0.05 بين قيمتي (معامل الارتباط بين معلمة الصعوبة الحقيقية وبين معلمة الصعوبة المقدرة للنموذج الثنائي وفق طريقة التقدير MML) من جهة و(معامل الارتباط بين معلمة الصعوبة الحقيقية وبين معلمة الصعوبة المقدرة للنموذج الثنائي وفق طريقة التقدير Bayes) من جهة أخرى عندما كان عدد الفقرات 10 فقرة وباختلاف حجم العينة (1000، 1500) لصالح (معامل الارتباط بين معلمة الصعوبة الحقيقية وبين معلمة الصعوبة المقدرة للنموذج الأحادي وفق طريقة التقدير Bayes).

وأظهرت النتائج وجود فرق دال إحصائياً عند مستوى الدلالة α=0.05 بين قيمتي (معامل الارتباط بين معلمة الصعوبة الحقيقية وبين معلمة الصعوبة المقدرة للنموذج الثنائي وفق طريقة التقدير MML) من جهة و(معامل الارتباط بين معلمة الصعوبة الحقيقية وبين معلمة الصعوبة المقدرة للنموذج الثنائي وفق طريقة التقدير Bayes) من جهة أخرى عندما كان عدد الفقرات 10 فقرة وباختلاف حجم العينة 500 فرداً لصالح (معامل الارتباط بين معلمة الصعوبة الحقيقية وبين معلمة الصعوبة المقدرة للنموذج الثنائي وفق طريقة التقدير MML). في حين لم يثبت وجود فرق دال إحصائياً عند مستوى الدلالة α=0.05 بين قيمتي (معامل الارتباط بين معلمة الصعوبة الحقيقية وبين معلمة الصعوبة المقدرة للنموذج الثنائي وفق طريقة التقدير MML) من جهة و(معامل الارتباط بين معلمة الصعوبة الحقيقية وبين معلمة الصعوبة المقدرة للنموذج الثنائي وفق طريقة التقدير Bayes) من جهة أخرى عندما كان عدد الفقرات 10 فقرات وحجم العينة 250 فرداً.

الفصل الخامس

مناقشة النتائج والتوصيات الخاصة بها

يتناول هذا الفصل عرضاً لنتائج الدراسة مبوبة حسب أسئلتها:

أولاً. النتائج المتعلقة بالإجابة عن سؤال الدراسة الأول:

"ما أثر طريقة التقدير (MML، وBayes) في دقة تقدير معالم الفقرة في ضوء اختلاف المتغيرات (النموذج اللوغاريتمي المستخدم وعدد الفقرات، وحجم العينة)؟"

أظهرت نتائج سؤال الدراسة الأول أن كافة قيم مؤشر RMSE كمؤشر دقة لتقديرات معلمة التمييز للنموذج الثنائي حسب طريقة التقدير MML باختلاف عدد الفقرات (10، 40) واختلاف حجم العينة (250، 500، 1000، 1500) قد كانت أكبر ظاهرياً من نظيراتها حسب طريقة التقدير Bayes باستثناء أن قيمة مؤشر RMSE كمؤشر دقة لتقديرات معلمة التمييز للنموذج الثنائي حسب طريقة التقدير MML عندما كان عدد الفقرات 10 فقرات وحجم عينة 250 قد كانت أصغر ظاهرياً من نظيرتها حسب طريقة التقدير بييز (Bayes)؛ مما يُعني أن طريقة التقديرBayes قد كانت أكثر دقة ظاهرياً في تقدير معلمة التمييز في النموذج (2-PL) مقارنة مع نظيرتها طريقة الأرجحية العظمى الهامشية (MML)، وجاءت قيم المؤشرRMSE متفقة مع النتائج الخاصة بتقديرات معالم الفقرة حول انخفاض قيم الخطأ المعياري لقيم معلمة التمييز المقدرة وفق طريقة التقدير Bayes، مما يدل على أفضلية طريقة التقدير Bayes، ويُعزى ذلك إلى أثر مراعاة التوزيعات السابقة لمعالم الفقرات في تقديرات الاحتمالية لعملية التقدير. وتتفق هذه النتائج مع دراسة هسيه وآخرين (Hsieh, Proctor, Hou & Teo, 2010) ، دي لا توري ويوان (De la Torre, Jimmy, Yuan Hong, 2010) التي أشارت إلى أن تقديرات معلمة التمييز وفق طريقة التقدير Bayes أكثر دقة من نظيرتها طريقة MML واتفقت النتائج مع دراسة لورد (Lord, 1986) ، بوزيهي (Pozeih, B.J, 1990)، (الدرابيع، وعليان، 2000)، ديمارس (Demars,C. E, 2001) (عبابنة، 2007)،

شولتز ويو(Liu, Schulz & Yu, 2009)، (الشريفين، 2012) ، ودراسة كل من ليو، ودراسة غاري وفيرمونت (Garee & Vermunt, 2006) ودراسة ليندن (linden, 1998) وأشارت جميع الدراسات السابقة إلى أفضلية طريقة بييز في تقدير معالم الفقرة، وانسجمت نتائج الجدول (4) مع دراستي عليان والدرابيع (2000) حول أثر زيادة عدد الفقرات على دقة تقدير معالم الفقرة والأفراد.

كما أظهرت نتائج سؤال الدراسة الأول أن كافة قيم مؤشر RMSE كمؤشر دقة لتقديرات معلمة الصعوبة للنموذج الأحادي حسب طريقة التقدير MML باختلاف عدد الفقرات (10، 40) واختلاف حجم العينة (250، 500، 1000، 1500) قد كانت **أصغر** ظاهرياً من نظيراتها حسب طريقة التقدير Bayes؛ مما يُعني أن طريقة MML كانت أكثر دقة ظاهرياً في تقدير معلمة الصعوبة من نظيرتها طريقة Bayes في تقدير معلمة الصعوبة، ويعزو الباحث هذه النتيجة إلى أنه للوهلة الأولى يظن بأنها نتيجة إيجابية تصب في صالح طريقة التقدير MML لكن سرعان ما تتداعى هذه النتيجة إذ تم التركيز على أنها مقترنة بالنموذج اللوغاريتمي الأحادي الذي يشتمل ضمنياً معلمة التمييز كعامل دخيل في تقديرات معلمة الصعوبة للنموذج اللوغاريتمي الأحادي، مما يشكك بموثوقية كافة النتائج الخاصة بطريقة التقدير MML في حالة النموذج اللوغاريتمي الأحادي.

وكذلك أظهرت نتائج سؤال الدراسة الأول أن كافة قيم مؤشر RMSE كمؤشر دقة لتقـديرات معلمة الصعوبة للنموذج الثنائي حسب طريقة التقدير MML باختلاف عـدد الفقـرات (10، 40) واختلاف حجم العينة (250، 500، 1000، 1500) قد كانت **أكبر** ظاهرياً من نظيراتهـا حسـب طريقة التقدير Bayes؛ مما يُعني أن طريقة Bayes أكثر دقة في تقدير معلمة الصعوبة مقارنـة بطريقةMML باختلاف عدد الفقرات وحجم العينة ويذهب الباحث في معرض تفسيره لهذه النتيجة

إلى أنها تتسم بموضوعية لا يستهان بها إذ تم تشخيص العامل الدخيل في قيم معلمة الصعوبة المقدرة وفقاً للنموذج اللوغاريتمي الأحادي المعلمة على هيئة معلمة التمييز في النموذج اللوغاريتمي الثنائي ولهذا تم إنصاف طريقة التقدير Bayes من حيث أنها لديها القدرة على تقدير معالم الفقرات (الصعوبة، والتمييز) بدقة أعلى من دقة طريقة التقدير MML في حالة النموذج اللوغاريتمي الثنائي. كما يُعزى ذلك إلى أثر مراعاة التوزيعات السابقة لمعالم الفقرات في تقديرات الاحتمالية وفق طريقة Bayes وخاصة عند تقدير معالم الفقرات لحجوم العينات الصغيرة، وعند عدد الفقرات الاختبارية القليلة، وجاءت نتائج قيم مؤشر RMSE متفقه مع ما أشار إليه لورد (Lord, 1986)، غولدمان وراجو (Goldman & Raju, 1986) ولينـدن (linden,1998) وساومينثان وجيفورد (Swaminathan & Gifford, 1982)، فيتزباترك وآن (Fitzpatrick & Ann, 2001)، وغاري وفيرمونت (Garee & Vermunt ,2006) وكل من هيونغ ولين وشين ,Huang ,Lin & Shin) (2001 بيرجوس (Burgos, 2010) في أن طريقة Bayes تعطي تقديرات أكثر دقة لمعالم الفقرات مقارنة بطريقة MML.

النتائج المتعلقة بالإجابة عن سؤال الدراسة الثاني:

"ما أثر طريقة التقدير (MML، وBayes) في الكفاءة النسبية لتقدير معالم الفقرة في ضوء اختلاف المتغيرات (النموذج اللوغاريتمي المستخدم وعدد الفقرات، وحجم العينة)؟"

أشارت نتائج سؤال الدراسة الثاني إلى أن قيم الكفاءة النسبية لطريقة Bayes على MML لتقديرات معالم الصعوبة في النموذج اللوغاريتمي الأحادي قد كانت ظاهرياً لصالح طريقة التقدير MML عندما يكون عدد الفقرات 10 على اختلاف حجوم العينات في النموذج اللوغاريتمي الأحادي؛ ويعزو الباحث هذه النتيجة إلى أنها مشكك فيها في ضوء أن عدد الفقرات قد كان أقل من عشرين فقرة مما يؤثر سلباً على دقة تقدير معالم الفقرة (الصعوبة) في حالة النموذج اللوغاريتمي

الأحادي، وأن قيم الكفاءة النسبية لطريقة Bayes على MML لتقديرات معالم الصعوبة في النموذج اللوغاريتمي الأحادي قد كانت ظاهرياً لصالح طريقة التقدير MML عندما يكون عدد الفقرات 40 في حجوم العينات (1000، 1500) في النموذج اللوغاريتمي الأحادي ويعزو الباحث هذه النتيجة إلى أنها تتفق مع الإطار النظري من حيث أنه كلما زاد عدد الفقرات وزادت حجوم العينات تزداد كفاءة التقدير لمعالم الفقرات وهذه النتيجة تتفق مع ما أشار إليه (Hambleton & Cook, 1983)، وعلى العكس من ذلك فإن قيم الكفاءة النسبية لطريقة Bayes على MML لتقديرات معالم الصعوبة في النموذج اللوغاريتمي الأحادي قد كانت ظاهرياً لصالح طريقة التقدير Bayes عندما يكون عدد الفقرات 40 في حجوم العينات (250، 500) في النموذج اللوغاريتمي الأحادي، ويعزو الباحث هذه النتيجة إلى مراعاة طريقة التقدير Bayes للتوزيعات السابقة أثناء عملية تقدير معالم الفقرات مما عاد بالإيجاب على الكفاءة النسبية لها. كما أشارت إلى أن قيم الكفاءة النسبية لطريقة Bayes على MML قد كانت ظاهرياً لصالح طريقة التقدير Bayes عندما يكون عدد الفقرات (10، 40) على اختلاف حجوم العينات في النموذج اللوغاريتمي الثنائي، ويعزو الباحث هذه النتيجة إلى مراعاة طريقة التقدير Bayes للتوزيعات السابقة أثناء عملية تقدير معالم الفقرات مما عاد بالإيجاب على الكفاءة النسبية لها.

كما أظهرت نتائج سؤال الدراسة الثاني أن قيم الكفاءة النسبية لطريقة Bayes على MML قد كانت ظاهرياً لصالح طريقة التقدير MML عندما يكون عدد الفقرات (10، 40) على اختلاف حجوم العينات في النموذج اللوغاريتمي الثنائي، ويعزو الباحث هذه النتيجة عندما يكون عدد الفقرات 10 فقرة إلى أنها مشكك فيها في ضوء أن عدد الفقرات قد كان أقل من عشرين فقرة مما يؤثر سلباً على دقة تقدير معالم الفقرة (الصعوبة) في حالة النموذج اللوغاريتمي الأحادي ومعالم الفقرة (التمييز) في حالة النموذج اللوغاريتمي الثنائي. ويعزو الباحث هذه النتيجة عندما يكون عدد

الفقرات 40 فقرة إلى مراعاة طريقة التقدير Bayes للتوزيعات السابقة أثناء عملية تقدير معالم الفقرات مما عاد بالإيجاب على الكفاءة النسبية لها وتتفق النتائج مع ما ذهب إليه كلٍ من غولدمان وراجو (Goldman & Raju, 1986)، فيتزباترك وآن (Fitzpatrick & Ann, 2001)، في الأثر الإيجابي لزيادة حجم العينة على كفاءة التقدير.

النتائج المتعلقة بالإجابة عن سؤال الدراسة الثالث:

"هل توجد فروق دالة إحصائيا عند مستوى الدلالة (α=0. 05) في قيم معاملات الارتباط بين معالم الفقرة الحقيقية وبين معالم الفقرة المقدرة وفق طريقتي التقدير (MML, Bayes) باختلاف المتغيرات (النموذج اللوغاريتمي المستخدم، عدد الفقرات، حجم العينة)؟"

أظهرت نتائج سؤال الدراسة الثالث وجود فرق دال إحصائياً عند مستوى الدلالة α=0.05 عندما كان عدد الفقرات 10 بين قيمتي (معامل الارتباط بين معلمة التمييز الحقيقية وبين معلمة التمييز المقدرة للنموذج الثنائي وفق طريقة التقدير MML) من جهة و(معامل الارتباط بين معلمة التمييز الحقيقية وبين معلمة التمييز المقدرة للنموذج الثنائي وفق طريقة التقدير Bayes) من جهة أخرى عندما كان عدد الفقرات 40 فقرة وباختلاف حجم العينة (250، 500، 1000، 1500) لصالح (معامل الارتباط بين معلمة التمييز الحقيقية وبين معلمة التمييز المقدرة للنموذج الثنائي وفق طريقة التقدير MML). كما أظهرت نتائج سؤال الدراسة الثالث وجود فرق دال إحصائياً عند مستوى الدلالة α=0.05 بين قيمتي (معامل الارتباط بين معلمة التمييز الحقيقية وبين معلمة التمييز المقدرة للنموذج الثنائي وفق طريقة التقدير MML) من جهة و(معامل الارتباط بين معلمة التمييز الحقيقية وبين معلمة التمييز المقدرة للنموذج الثنائي وفق طريقة التقدير Bayes) من جهة أخرى عندما كان عدد الفقرات 10 فقرة وباختلاف حجم العينة (500، 1000، 1500) لصالح (معامل

الارتباط بين معلمة التمييز الحقيقية وبين معلمة التمييز المقدرة للنموذج الثنائي وفق طريقة التقـدير Bayes). وكذلك أظهرت نتائج سؤال الدراسة الثالث أنه لم يثبت وجود فرق دال إحصـائياً عنـد مستوى الدلالة α=0.05 بين قيمتي (معامل الارتباط بين معلمة التمييز الحقيقية وبين معلمة التمييز المقدرة للنموذج الثنائي وفق طريقة التقدير MML) من جهة و(معامل الارتباط بين معلمة التمييـز الحقيقية وبين معلمة التمييز المقدرة للنموذج الثنائي وفق طريقة التقدير Bayes) من جهة أخـرى عندما كان عدد الفقرات 10 فقرة وحجم العينة 250 فرداً.

كذلك أظهرت نتائج سؤال الدراسة الثالث وجود فرق دال إحصـائياً عنـد مسـتوى الدلالـة α=0.05 بين قيمتي (معامل الارتباط بين معلمة الصعوبة الحقيقية وبين معلمة الصعوبة المقـدرة للنموذج الأحادي وفق طريقة التقدير MML) من جهة و(معامل الارتباط بـين معلمـة الصـعوبة الحقيقية وبين معلمة الصعوبة المقدرة للنموذج الأحادي وفق طريقة التقـدير Bayes) مـن جهـة أخرى عندما كان عدد الفقرات 10 فقرة وباختلاف حجـم العينـة (250، 500، 1000، 1500) لصالح (معامل الارتباط بين معلمة الصعوبة الحقيقية وبين معلمة الصعوبة المقدرة للنموذج الأحادي وفق طريقة التقدير Bayes). كما أظهرت نتائج سؤال الدراسة الثالث وجود فرق دال إحصائياً عند مستوى الدلالة α=0.05 بين قيمتي (معامل الارتباط بين معلمة الصعوبة الحقيقية وبـين معلمـة الصعوبة المقدرة للنموذج الأحادي وفق طريقة التقدير MML) من جهة و(معامل الارتبـاط بـين معلمة الصعوبة الحقيقية وبين معلمة الصعوبة المقدرة للنموذج الأحـادي وفـق طريقـة التقـدير Bayes) من جهة أخرى عندما كان عدد الفقرات 40 فقرة وباختلاف حجم العينة (500، 1000، 1500) لصالح (معامل الارتباط بين معلمة الصعوبة الحقيقية وبين معلمة الصعوبة المقدرة للنموذج الأحادي وفق طريقة التقدير Bayes)، وكذلك أظهرت نتائج سؤال الدراسة الثالث أنه لـم يثبـت وجود فرق دال إحصائياً عند مستوى الدلالة α=0.05 بين قيمتي (معامل الارتبـاط بـين معلمـة

الصعوبة الحقيقية وبين معلمة الصعوبة المقدرة للنموذج الأحادي وفق طريقة التقدير MML) مــن جهة و(معامل الارتباط بين معلمة الصعوبة الحقيقية وبين معلمة الصعوبة المقدرة للنموذج الأحادي وفق طريقة التقدير Bayes) من جهة أخرى عندما كان عدد الفقرات 40 فقرة وحجم العينة 250 فرداً.

ويُعزي الباحث ذلك إلى أثر مراعاة التوزيعات السابقة وفق طريقة Bayes في دقة تقدير معالم الفقرات، وخاصة عند تقدير معالم الفقرات عند حجوم العينات الصغيرة، وعند عدد الفقرات الاختبارية القليلة، وهذه النتيجة متفقة مع ما أشار إليه لورد (Lord, 1986) وليندن (linden, 1998) وساومينثان وجيفورد (Swaminathan & Gifford, 1982) وغاري وفيرمونت (Garee & Vermunt, 2006) وهيونغ ولين وشين (Huang, Lin & Shin, 2001) جيو وشين (Geo & Cheen, 2005) في أن طريقة Bayes تعطي تقديرات أكثر دقة لمعالم الفقرات مقارنة مع طريقة MML. في حين لم تتفق النتائج مع ما توصلت له دراسة نيرنج (Nerng, 1995)، بارنيز ووايز (Barnes & Wise, 1991).

كذلك أظهرت نتائج سؤال الدراسة الثالث وجود فرق دال إحصائياً عند مستوى الدلالة α=0.05 بين قيمتي (معامل الارتباط بين معلمة الصعوبة الحقيقية وبين معلمة الصعوبة المقدرة للنموذج الثنائي وفق طريقة التقدير MML) من جهة و(معامل الارتباط بين معلمة الصعوبة الحقيقية وبين معلمة الصعوبة المقدرة للنموذج الثنائي وفق طريقة التقدير Bayes، من جهة أخرى عندما كان عدد الفقرات 40 فقرة وباختلاف حجم العينة (250، 1500) لصالح (معامل الارتباط بين معلمة الصعوبة الحقيقية وبين معلمة الصعوبة المقدرة للنموذج الأحادي وفق طريقة التقدير Bayes). في حين لم تظهر نتائج سؤال الدراسة وجود فرق دال إحصائياً عند مستوى الدلالة α=0.05 بين قيمتي (معامل الارتباط بين معلمة الصعوبة الحقيقية وبين معلمة الصعوبة المقدرة

للنموذج الثنائي وفق طريقة التقدير MML) من جهة و(معامل الارتباط بين معلمة الصعوبة الحقيقية وبين معلمة الصعوبة المقدرة للنموذج الثنائي وفق طريقة التقدير Bayes) من جهة أخرى عندما كان عدد الفقرات 40 فقرة وحجم العينة (500، 1000) فردا. كذلك أظهرت نتائج سؤال الدراسة الثالث وجود فرق دال إحصائياً عند مستوى الدلالة α=0.05 بين قيمتي (معامل الارتباط بين معلمة الصعوبة الحقيقية وبين معلمة الصعوبة المقدرة للنموذج الثنائي وفق طريقة التقدير MML) من جهة و(معامل الارتباط بين معلمة الصعوبة الحقيقية وبين معلمة الصعوبة المقدرة للنموذج الثنائي وفق طريقة التقدير Bayes) من جهة أخرى عندما كان عدد الفقرات 10 فقرة وباختلاف حجم العينة (1000، 1500) لصالح (معامل الارتباط بين معلمة الصعوبة الحقيقية وبين معلمة الصعوبة المقدرة للنموذج الأحادي وفق طريقة التقدير Bayes)، ويُعزي الباحث ذلك إلى أثر مراعاة التوزيعات السابقة وفق طريقة Bayes في دقة تقدير معالم الفقرات، وعند عدد الفقرات الاختبارية القليلة، وهذه النتيجة متفقة مع ما أشار إليه لورد (Lord, 1986) وليندن (linden, 1998) وساوميناثان وجيفورد (Swaminathan & Gifford, 1982) وغاري وفيرمونت (Garee & Vermunt ,2006) وهيونغ ولين وشين(Huang, Lin & Shin, 2001) في أن طريقة Bayes تعطي تقديرات أكثر دقة لمعالم الفقرات مقارنة مع طريقة MML.

وكذلك أظهرت نتائج سؤال الدراسة الثالث وجود فرق دال إحصائياً عند مستوى الدلالة α=0.05 بين قيمتي (معامل الارتباط بين معلمة الصعوبة الحقيقية وبين معلمة الصعوبة المقدرة للنموذج الثنائي وفق طريقة التقدير MML) من جهة و(معامل الارتباط بين معلمة الصعوبة الحقيقية وبين معلمة الصعوبة المقدرة للنموذج الثنائي وفق طريقة التقدير Bayes) من جهة أخرى عندما كان عدد الفقرات 10 فقرة وباختلاف حجم العينة 500 فرداً لصالح (معامل الارتباط بين معلمة الصعوبة الحقيقية وبين معلمة الصعوبة المقدرة للنموذج الثنائي وفق طريقة التقدير MML)، وهذه

النتيجة كذلك بسبب انخفاض عدد الفقرات دون الــ 20 فقرة، فهي مشكك بها في ضوء المعلومات الخاصة بالاختبارات القصيرة دون الــ 20 فقرة في دليل تشغيل برنامج Bilog-MG v3. في حين لم تثبت نتائج سؤال الدراسة الثالث وجود فرق دال إحصائياً عند مستوى الدلالة α=0.05 بين قيمتي (معامل الارتباط بين معلمة الصعوبة الحقيقية وبين معلمة الصعوبة المقدرة للنموذج الثنائي وفق طريقة التقدير MML) من جهة و(معامل الارتباط بين معلمة الصعوبة الحقيقية وبين معلمة الصعوبة المقدرة للنموذج الثنائي وفق طريقة التقدير Bayes) من جهة أخرى عندما كان عدد الفقرات 10 فقرات وحجم العينة 250 فرداً، مما يعني أن طريقتي التقدير ذات دقة متقاربة عند حجوم العينات الصغيرة وعدد الفقرات القليلة ويعزي الباحث هذه النتيجة إلى عدم موثوقيتها حيث نتجت عن تقديرات MML التي يشكك بها في مثل هذه المواقف، وهذا ما أشار إلية دليل التشغيل لبرنامج Bilog (DU,Toil,M. 2003)، حول عدم مقدرة طريقة MML تقدير معالم الفقرات بدقة عند حجم عينة 250.

الاستنتاجات والتوصيات

أظهرت نتائج الدراسة أن طريقة تقدير معالم الفقرة باستخدام طريقة Bayes قد كانت أكثر دقة من طريقة MML في معظم مواقف الدراسة باختلاف المتغيرات (النموذج، عدد الفقرات، حجم العينة)، وقد يعزى ذلك إلى أسباب عدة.

1. إن طريقة Bayes في التقدير تعتمد على إجراءات أكثر ضبطا من نظيرتها (طريقة MML)، حيث إن هذه الطريقة تعتمد على التوزيعات السابقة (Prior Distribution) عند عملية تقدير الاحتمالية، للوصول إلى أعلى دقة لمعالم الفقرة، مقارنة مع نظيرتها طريقة MML التي تعتمد توزيعا ثابتا عند عملية تقدير الاحتمالية، لمعالم الفقرة، لذلك أعطت طريقة Bayes دقة أكبر في تقدير معالم الفقرة.
2. عند استخدام حجوم عينات صغيرة وعدد فقرات قليلة تصبح تقديرات معالم الفقرة وفق طريقة MML غير موثوقة وتفتقر للدقة، لذلك تكون طريقة Bayes مناسبة لمعايرة معالم الفقرات في مثل هذه المواقف.
3. أظهرت الدراسة تقارباً بمعيار الدقة بين القيم المقدر وفق الطريقتين (MML, Bayes) عند حجوم عينات كبيرة، وغالباً لا يمكن فصلهما، مما يؤكد مبدأ التكافؤ بينهما، لذلك يمكن استخدام أحدهما لتأكيد نتيجة الأخرى.
4. تم توليد البيانات (الحقيقية) لمعالم الفقرات باستخدام برنامج (WinGen v. 3) تحت افتراض التوزيع الطبيعي بوسط (0) وانحراف معياري (1) لمعلمة الصعوبة، أما معلمة التمييز فكانت قيمتها الابتدائية (0. 4) وقيمتها النهائية (1. 7)، وبناءً على القيم المقدرة لمعالم الفقرات أظهرت نتائج الدراسة أن طريقة Bayes تعطي دقة أعلى في تقدير معالم الفقرة في معظم مواقف الدراسة، وفق التوزيعات سابقة الذكر.

5. إن البرنامج الحاسوبي (WinBUGS V 1. 4) المستخدم في هذه الدراسة، الذي تم تطـويره من قبل مجلس البحوث الطبية وحدة الإحصاء الحيوي في كامبريدج، المملكة المتحدة، يعتبر من البرامج النادرة التي تعمل وفق منحى الاستدلال البييزي، قد يساعد إذا أحسن اسـتخدامه فـي الوصول إلى معالم فقرات أكثر دقة من البرامج الحاسوبية الأخرى وخاصة برنامج Bilog-MG3 الذي يعمل بمنحى استدلالي يقوم على طريقة الأرجحية العظمى الهامشية.

وبناءًا على نتائج الدراسة يوصي الباحث بما يأتي:

1. إجراء دراسة مماثلة على بيانات واقعية، من الميدان التربوي يمكن الحصول من خلالها علـى التوزيعات السابقة لمعالم الفقرات.
2. إجراء دراسة مماثلة على النموذج اللوغاريتمي ثلاثي المعلمة.
3. إعادة الدراسة على نماذج نظرية استجابة الفقرة للاستجابات المتدرجة؛ للوقوف على أي مـن الطريقتين يناسب هذا النوع من الاختبارات.
4. تفعيل منحى بييز في معايرة Calibration فقرات بنوك الأسـئلة، للاسـتفادة مـن خاصـية التوزيعات السابقة (Prior Distribution) التي تزيد من دقة تقدير معالم الفقرات والأفراد.

المراجع العربية:

أبوعلام، رجاء. (2005). **تقويم التعليم**، عمان: دار الميسرة.

التقي، أحمد. (2010). **النظرية الحديثة في القياس**، عمان: دار الميسرة.

الثوابية، أحمد. (2010). أثر حجم العينة على تقدير صعوبة الفقرة والخطأ المعياري في تقديرها باستخدام نظرية الاستجابة للفقرة. **مجلة جامعة دمشق**، 26(2) 525-556

الدرابيع، ماهر. (2001). فعالية النموذج اللوغاريتمي ذي المعلمة الواحدة "نموذج راش" في دقة تقدير قدرة الفرد ومعامل صعوبة الفقرة باختلاف حجم العينة وطول الاختبار. **مجلة دراسات – العلوم الانسانية**، 28(1)، 197-208.

الشريفين، نضال. (2012). أثر طريقة تقدير معالم الفقرة وقدرات الأفراد على قيم معالم الفقرة والخصائص السيكومترية للاختبار، في ضوء تغير حجم العينة. **المجلة التربوية**، 46 (104) 177-238.

عبابنة، عماد. (2004). **أثر حجم العينة وطريقة انتقائها وعدد الفقرات وطريقة انتقائها على دقة تقدير معالم الفقرة والقدرة لاختبار قدرة عقلية باستخدام نظرية الاستجابة للفقرة.** رسالة دكتوراه، جامعة عمان العربية، عمان، الاردن.

عبابنة، عماد. (2006). مقارنة فاعلية طريقة الأرجحية العظمى وطريقة بييز في تقدير معلمة القدرة عند استخدام النموذج اللوغاريتمي الثلاثي، **مجلة الاكاديمية العربية المفتوحة**، الدنمارك، (3) 5-22.

علام، صلاح الدين. (2005). **نماذج الاستجابة للمفردة الاختبارية أحادية البعد ومتعددة الأبعاد وتطبيقاتها في القياس النفسي والتربوي.** القاهرة: دار الفكر العربي.

علام، صلاح الدين. (2006). **القياس والتقويم التربوي النفسي، أساسياته وتطبيقاته وتوجهاته المعاصرة**. القاهرة: دار الفكر العربي.

عليان، خليل والدرابيع، ماهر. (2000). أثر حجم العينة وطول الاختبار والتفاعل بينهما على دقة تقدير قدرة الفرد ومعالم الفقرة عند استخدام النموذج اللوغاريتمي ذي الثلاثة معالم، **مجلة دراسات**، الجامعة الاردنية، 27 (2) 464 - 471

عودة، أحمد، والخليلي، خليل. (2000). **الإحصاء للباحث في التربية والعلوم الإنسانية**. الاردن: دار الأمل.

عوده، أحمد. (2010). **القياس والتقويم في العملية التدريسية**، الاردن: دار الامل.

المراجع الاجنبية :

Baker, Frank B. (2001). **The Basic of Item Response Theory**. ERIC Clearinghouse on Assessment an Evaluation.

Barners, L. L. & Wise, S. L. (1991):The Utility of a Modified one-parameter IRT Model with small Samples. **Applied Measurement in Education,** 4 (2), 143-157.

Bellhouse, D.R. (2004). The Reverend Tomes Bayes, FRS: A Biography to celebrate the Tercentenary of His Birth. **Statistical Science** 19 (1), 3-43.

Borsboon, D . Mellenbergh, G, & Heerden, J.V. (2002). Defferent Kinds of DIF A Distinction Between Absolute and Relatve Forms of Measurement Invariance and Bias. **Applied Psychological Measurement,** 26 (4),433-350.

Burgos, J. G. (2010). Bayesian Mothods in Psychological Research. The case of IRT. **Internationl Journal of Psychological Research,** 3(1), 164- 176.

Christian, P. George, C. (2008). **Monte Carlo Statistical Methods.** Springer texts in Statistics, Second Edition.

Cowles, K. (2004). "Review of WinBuGs 1.4" **The American statistician,** 58, 330- 336.

Crocker, L. & Algina, J. (1986). **Introduction to Classical and Modern Test Theory**. New York , CBS College Publishing.

De la Torre, Jimmy, Yuan Hong(2010). Parameter Estimation With Small Sample Size A Higher-Order IRT Model Approach. **Applied Psychological Measurement.** . 34 (4), 267-285.

Demars, C. E. (2001). Group Differences based on IRT Scores:dose the Model matter?. **Educational and Psycological Measurement,**60 (1),60-70

Du, Toit, M. (2003). **IRT from SSI: BILOG-MG, MULTILOG, PARSCALE, TEST-FACT.** Lincolnwood, IL: Scientific Software International.

Embretson, S, E, & Reise, S. P. (2000). **Item Respouse Theory For Psychologistis,** Lawrennce Erlbaum Associates, new Jersey Publishers Mahwah, London.

Ferrando, P. J. , Lorenzo - Seva, U. , & Molina, G. (2001). An item response theory analysis of response stability in personality measurement. **Applied Psychological Measurement.** 25(1), 3-19.

Fitzpatrick, Ann. R. (2001). The effects of test length and sample size on the reliability and equating of tests Composed of constructed -response items. **Applied Measurement in Education** ,14,issue 1.

Frank, Rijmen. (2009). **Efficient full in Formation Maximum Likelihood Estimation for Multidiman Sional IRT Models,** Educational Testing service (ETS).

Gao, F. & Chen, L. (2005). Bayesian or non-Bayesian: A comparison study of itemparameter estimation in the three-parameter logistic model. **Applied Measurement in Education,** 18, 351-380.

Garre. G. & Vermunt. K. (2006). Avoiding Boundary Estimation in Latent Class Analysis by Bayesian Posterior Mode Estimation. **Behaviormentrika** 33. (1), 256-271.

Glas, C. A, W., & Falcon, J. C. S. (2003). A comparison of item – fit statistics for the three–parameter logisticmodel. **Applied Psychological Measurement,** 27 (2), 89- 106.

Goldman , S. H & Raju , N. S, (1986). Recovery of One and Tow parameter Logistic Item parameters: an Empirical Study. **Educational and Psychological Measurement,** 46 (1),11-21.

Hambleton, R. K, Cook. L. L. (1983). Robustness of Item Response Models and Effects of Test Length and Sample Size in the Precision of Ability Estimates. New York. In D. J. Weiss (Ed), **New Horizons in Testing,** 31-49.

Hambleton, R. K. & Swaminathan, H. (1985). **Item Response Theory: Principles and Applications,** Boston, Kluwer, Nijhoff Publishing.

Hambleton, R. K., & Cook, L. L. (1977). Latent trait models and their use in analyzing educational test data. **Journal of Educational Measurement,** 14(2), 75-96.

Hattie, J. (1985). Methodology Review: Assessing unidmensionality of tests and item. **Applied Psychological Measurement,** 9 (2),139-164.

Hsieh, N., Proctor, T. , Hou, J., & Teo, K. (2010). A comparison of Bayesian MCMC and Marginal Maximum Likelihood Methods in Estimationg the Item Prameters for the 2Pl 1RT Model, **internetionl Journal of Innovative Management, Information & production,** 1 (1), 81- 89.

Huang, C. Y, Lohss. W. E , Lin ,& shin, D. (2001). Item calibration of licensure test with multiple specialty components. Submitted to Division DI: **Educational Measurement, Psychometrics and assessment.** Enabled Tiger ,a web – based Manuscripts Processing system. Michigan state University.

Koning, E. D., Sijtsma, K. & Hamers, J. H. M. (2002). Comparion of Four IRT Modele When Analyzing Two Test For Inductive Reasoning. **Applied Psychological Measurement,** 26 (3), 302-320.

Linden,W. (1998). Bayesian Item Selection Criteria of Adaptive Testing. **Psychometrika,** 63(2)201-216.

Liu, Y. M. , Schulz, E. M. , & Yu, L. (2009). Standard error estimation of 3-PL IRT true score equating with an MCMC method. **Journal of Educational and Behavioral Statistics,** 33,257-278.

Lo, S., Liu, X. & Shao,Y. (2003).A Marginal likelihood for Family Bata . **Annals of Human Genetics**, 63,357-366.

Lord, M. F. (1980) **Application of Item Response Theory to Practical Testing Problems**. Hillsdale, NJ: Erlbaum.

Lord, M. F. (1986). Maximum Likelihood, and Bayesian parameter Estimation in Item Response Theory. **Journal of Educational Measurement**, 23 (2),157-162.

Nerng, M. L. (1995). The distribution of person fit Using true and estimated Person Parameters. **Applied Psychological Measurement**, 19, 121-129.

Pozeih, B. J. (1990). **Application of item Response Theory to Criterion - Referenced Measurement: An Investigation of the effects of Model Choice, Sample size, and test length on Reliability and Estimation Accuracy.** PH.D university of Nebraska, Lincoln, Nebraska, Lincoln, May.

Swaminathan, H., & Gifford, J. , A. (1982). Bayesian in the Rasch Model. **Journal of Educational Statistics**, 7 (3), 175- 191.

Sykes, R. C. & Hou, L. (2003).Weighting Constructed-Response Item in IRT-Based Exams . **Applied Measurement in Education** ,16(4),257-275.

Zimowski, M. F., Muraki, E., Mislevy, R., J., & Bock, R. D. (2003). **Biilog, Mg3** (Computer software).

الملاحق

الملحق (1)

ملف أوامر برنامج Bilog للنموذج أحادي المعلمة Rash عندما يكون عدد الفقرات (10) وعدد الأفراد (250) فرد.

```
>GLOBAL DFName = 'C:\SS\1pl\10i\250p\Bilog\run1\250p10i. wgr',
         NPArm = 1,
         LOGistic,
         SAVe;
>SAVE PARm = 'C:\SS\1pl\10i\250p\Bilog\run1\run11pl10i250p.
PAR',
      SCOre = 'C:\SS\1pl\10i\250p\Bilog\run1\run11pl10i250p.
SCO';
>LENGTH NITems = (10);
>INPUT NTOtal = 10,
       NALt = 2,
       NIDchar = 10;
>ITEMS ;
>TEST1 TNAme = 'TEST0001',
       INUmber = (1(1)10);
(10A1, 10A1)
>CALIB CYCles = 400,
       NEWton = 0,
       PLOt = 0. 0100,
       ACCel = 1. 0000,
       EMPirical,
       TPRior,
       NOSprior,
       FLOat,
       CHIsquare = (10, 9),
       RASch;
>SCORE METhod = 1,
       INFo = 2,
       FIT;
```

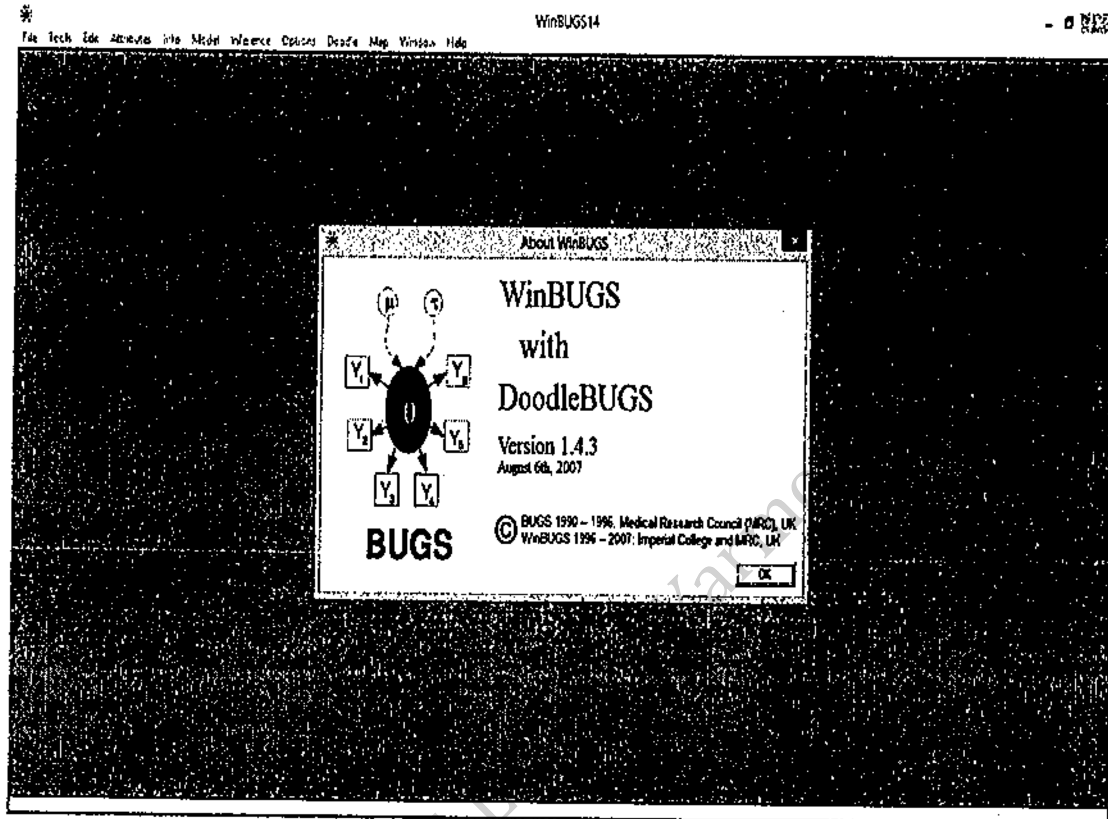
الملحق (2)

ملف أوامر من برنامج Bilog للنموذج ثنائي المعلمة عندما يكون عدد الفقرات 10 وعدد الأفراد 500 فرداً

```
>GLOBAL DFName = 'C:\SS\2-PL \10i\500p\Bilog\run1\500p10i. wgr',
        NPArm = 2,
        LOGistic,
        SAVe;
>SAVE PARm = 'C:\SS\2-PL \10i\500p\Bilog\run1\run12-PL 10i500p.
PAR',
      SCOre = 'C:\SS\2-PL \10i\500p\Bilog\run1\run12-PL 10i500p.
SCO';
>LENGTH NITems = (10);
>INPUT NTOtal = 10,
       NALt = 2,
       NIDchar = 10;
>ITEMS ;
>TEST1 TNAme = 'TEST0001',
       INUmber = (1(1)10);
(10A1, 10A1)
>CALIB CYCles = 400,
       NEWton = 0,
       PLOt = 0. 0100,
       ACCel = 1. 0000,
       EMPirical,
       TPRior,
       NOGprior,
       FLOat,
       CHIsquare = (10, 9);
>SCORE METhod = 1,
       INFo = 2,
       FIT;
```

الملحق (3)

تعريف بواجهة برنامج WinBugs:



الملحق (4)

ملف أوامر مستمد من برنامج WinBugs عندما يكون النموذج أحادي المعلمة وعدد الفقرات 10 وحجم العينة 250 فرداً.

```
Model
}

# Read in individual item responses
for (i in 1:250)
}
for (j in 1:10)
}
x[i,j] <- response[i,j]
{
{

#Identify two-parameter logistic (1-PL) model
for (i in 1:250)
}
for (j in 1:10)
}
p[i,j] <- exp((theta[i]-b[j]))/(1+exp((theta[i]-b[j])))
x[i,j] ~ dbern(p[i,j])
mx[i,j] <- 1-x[i,j]
{

#Specify prior for examinee parameters
theta[i] ~ dnorm(0,1)
{

#Specify priors for item parameters
for (j in 1:10)
}
a[j] ~ dlnorm(1,1. 649)
b[j] ~ dnorm(0,2)
{
#Generate replicate data for posterior predictive checks
for (i in 251:3000)
}
for (j in 1:10)
}
p[i,j] <- exp((theta[i]-b[j]))/(1+exp((theta[i]-b[j])))
r[i,j] ~ dbern(p[i,j])
mr[i,j] <- 1-r[i,j]
{
theta[i] ~ dnorm(0,1)
{
#compute odds ratios ofr relicate data
for (k in 2:10)
}
for (i in 1:(k-1))
}
or[k,i]<-
inprod(r[251:3000,i],r[251:3000,k])*inprod(mr[251:3000,i],mr[251:3000,k])/(inprod(mr[251:3000,i],r[251:3000,k])*inprod(r[2
51:3000,i],mr[251:3000,k]))
{
{

# End of model
{
```

الملحق (5)

توصيف ملف البيانات الذي يلحق في ملف الأوامر الخاصة به، وذلك على النحو الآتي:

```
# listing of item response data
list(response=structure(. Data=c(
1, 1, 1, 0, 1, 1, 1, 0, 0, 1,
0, 1, 1, 0, 1, 1, 0, 1, 1, 0,
1, 0, 1, 1, 1, 1, 1, 0, 1, 1,
0, 1, 0, 1, 0, 0, 0, 1, 1, 1,
0, 0, 1, 1, 0, 1, 1, 0, 1, 1,
0, 0, 1, 0, 0, 1, 1, 1, 1, 1,
0, 1, 1, 0, 1, 0, 0, 0, 1, 0,
0, 1, 1, 1, 0, 1, 1, 1, 1, 0,
0, 1, 1, 1, 0, 1, 1, 1, 0, 1,
0, 0, 1, 1, 1, 1, 0, 1, 1, 1,
0, 0, 1, 1, 0, 1, 0, 1, 1, 1,
1, 1, 0, 0, 1, 0, 0, 1, 0, 0,
1, 0, 1, 1, 1, 0, 1, 1, 1, 1,
0, 0, 0, 0, 1, 0, 0, 1, 0, 1,
0, 1, 1, 1, 0, 1, 0, 1, 0, 0,
1, 0, 1, 0, 1, 1, 0, 0, 0, 0,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
0, 0, 1, 1, 0, 0, 1, 0, 0, 0,
1, 1, 1, 1, 1, 0, 1, 0, 1, 0,
1, 0, 1, 1, 0, 1, 1, 1, 1, 1,
0, 0, 0, 1, 0, 1, 1, 1, 0, 1,
0, 1, 1, 1, 0, 1, 1, 1, 0, 1,
1, 0, 1, 1, 1, 1, 1, 1, 1, 1,
0, 0, 1, 0, 0, 1, 1, 0, 0, 0,
1, 1, 0, 1, 0, 0, 1, 1, 1, 1,
1, 0, 0, 0, 0, 1, 0, 1, 1, 1,
1, 0, 0, 1, 0, 1, 1, 0, 0, 1,
0, 1, 1, 1, 1, 1, 1, 1, 1, 0,
0, 1, 1, 1, 0, 1, 1, 0, 1, 0,
0, 1, 0, 1, 1, 0, 1, 1, 1, 1,
0, 1, 1, 1, 1, 1, 1, 1, 1, 1,
0, 1, 1, 1, 1, 0, 1, 1, 1, 1,
0, 0, 0, 0, 0, 0, 1, 0, 1, 1,
0, 0, 0, 0, 0, 1, 0, 1, 1, 0,
0, 1, 0, 1, 0, 1, 0, 1, 1, 0,
0, 0, 1, 1, 0, 0, 0, 1, 1, 0,
0, 0, 1, 1, 1, 1, 1, 1, 0, 1,
0, 0, 1, 0, 1, 0, 1, 1, 1, 1,
0, 0, 1, 1, 1, 0, 1, 1, 1, 1,
0, 1, 1, 1, 1, 0, 1, 0, 0, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
0, 0, 1, 1, 0, 0, 1, 1, 0, 1,
0, 0, 0, 1, 1, 1, 1, 0, 1, 0,
0, 0, 0, 0, 1, 0, 1, 1, 0, 0,
1, 0, 0, 0, 1, 1, 0, 1, 0, 1,
1, 0, 0, 1, 0, 0, 1, 1, 0, 0,
0, 0, 0, 1, 0, 0, 1, 1, 1, 0,
0, 1, 1, 1, 1, 1, 1, 1, 1, 1,
0, 0, 1, 1, 1, 1, 1, 1, 1, 1,
0, 1, 1, 1, 0, 0, 0, 0, 1, 1,
0, 0, 1, 1, 1, 1, 1, 1, 1, 1,
1, 0, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 0, 1, 0, 0, 0, 0, 1, 0,
1, 0, 0, 1, 0, 0, 1, 1, 0, 1,
0, 1, 0, 1, 0, 0, 0, 1, 1, 1,
1, 1, 0, 1, 0, 0, 1, 1, 1, 1,
0, 0, 0, 0, 1, 0, 1, 1, 1, 1,
0, 1, 0, 1, 1, 1, 0, 1, 1, 0,
1, 1, 1, 1, 0, 1, 1, 1, 1, 1,
0, 1, 1, 0, 1, 1, 1, 0, 1, 0,
0, 1, 1, 1, 0, 1, 1, 0, 1, 1,
0, 1, 0, 1, 1, 1, 1, 1, 0, 1,
0, 1, 1, 1, 1, 0, 0, 0, 1, 1,
0, 0, 1, 1, 0, 1, 0, 0, 0, 1,
0, 1, 1, 1, 0, 1, 1, 1, 0, 0,
0, 0, 1, 1, 0, 1, 1, 0, 1, 1,
0, 0, 1, 0, 0, 1, 0, 0, 0, 0,
0, 1, 1, 1, 1, 0, 1, 1, 0, 1,
0, 1, 0, 1, 1, 1, 1, 0, 1, 1,
1, 1, 1, 1, 0, 1, 1, 1, 0, 1,
1, 1, 1, 1, 1, 1, 1, 0, 0, 0,
```

0, 1, 1, 1, 1, 1, 1, 1, 1, 0,
1, 0, 1, 1, 0, 1, 1, 1, 1, 1,
0, 1, 1, 1, 0, 1, 1, 1, 1, 1,
1, 1, 0, 1, 0, 1, 1, 0, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 0, 0,
1, 0, 0, 0, 1, 0, 1, 1, 1, 1,
0, 1, 1, 1, 0, 1, 1, 1, 1, 0,
1, 1, 0, 0, 0, 1, 1, 1, 1, 1,
0, 1, 1, 1, 0, 1, 1, 1, 1, 0,
0, 0, 0, 0, 0, 1, 1, 0, 0, 0,
0, 1, 1, 1, 0, 0, 1, 1, 0, 1,
0, 1, 1, 1, 1, 0, 1, 0, 1, 0,
0, 0, 1, 1, 1, 1, 1, 1, 1, 1,
0, 1, 0, 1, 1, 1, 1, 0, 0, 0,
0, 1, 1, 0, 0, 0, 1, 0, 0, 1,
0, 1, 0, 1, 0, 0, 0, 0, 1, 0,
1, 0, 0, 0, 1, 0, 0, 1, 0, 0,
0, 1, 1, 1, 1, 0, 1, 1, 1, 1,
0, 0, 0, 0, 1, 1, 0, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
0, 1, 1, 1, 1, 0, 1, 1, 1, 1,
0, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 0, 1, 0, 0, 1, 0, 1, 0, 1,
0, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 0, 0, 0, 1, 0, 0, 0, 1, 0,
1, 1, 1, 0, 1, 1, 1, 0, 1, 0,
1, 1, 1, 1, 0, 1, 1, 1, 1, 1,
0, 0, 1, 1, 0, 1, 0, 1, 0, 0,
0, 1, 0, 1, 0, 1, 1, 0, 0, 0,
0, 1, 1, 1, 0, 1, 1, 0, 1, 0,
1, 1, 1, 1, 0, 1, 1, 0, 1, 1,
0, 1, 1, 1, 1, 0, 1, 1, 1, 1,
0, 1, 0, 0, 1, 0, 0, 0, 1, 0,
0, 0, 1, 1, 1, 1, 1, 1, 1, 1,
0, 1, 1, 1, 0, 1, 1, 1, 0, 0,
0, 1, 1, 0, 1, 1, 1, 1, 1, 1,
0, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 0, 1, 1, 1, 0, 0, 0, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
0, 1, 1, 1, 0, 1, 0, 0, 0, 1,
0, 0, 0, 0, 0, 0, 1, 0, 1, 1,
1, 0, 0, 0, 0, 0, 1, 1, 0, 0,
1, 0, 1, 1, 0, 1, 1, 1, 1, 1,
1, 0, 1, 1, 0, 0, 1, 0, 1, 1,
0, 1, 1, 1, 0, 1, 1, 1, 0, 1,
1, 1, 1, 1, 0, 1, 1, 0, 0, 1,
0, 1, 0, 1, 0, 1, 1, 1, 1, 1,
0, 1, 1, 1, 0, 1, 1, 0, 1, 0,
0, 1, 0, 0, 0, 0, 0, 0, 1, 1,
0, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 0, 1, 1, 1, 0, 1, 1, 1, 0,
1, 1, 0, 1, 0, 1, 0, 1, 1, 1,
0, 1, 0, 1, 0, 0, 1, 1, 1, 0,
1, 0, 0, 1, 0, 0, 1, 1, 1, 0,
0, 1, 1, 1, 1, 1, 1, 1, 0, 1,
0, 1, 0, 1, 1, 0, 0, 0, 1, 1,
0, 1, 1, 1, 0, 0, 1, 0, 1, 0,
1, 0, 1, 1, 1, 1, 0, 1, 1, 1,
0, 0, 0, 1, 1, 0, 1, 0, 0, 0,
1, 1, 1, 1, 1, 1, 1, 1, 1, 0,
0, 1, 1, 1, 0, 0, 1, 0, 0, 0,
0, 0, 0, 0, 1, 1, 1, 0, 1, 1,
1, 0, 1, 1, 0, 1, 1, 1, 0, 0,
0, 0, 1, 0, 1, 1, 1, 1, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 1, 1,
1, 0, 1, 1, 1, 1, 1, 1, 1, 0,
0, 0, 0, 1, 0, 0, 1, 1, 0, 0,
0, 1, 1, 1, 0, 0, 1, 1, 1, 1,
0, 0, 1, 1, 0, 0, 1, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 1, 1, 0,
0, 1, 0, 1, 1, 0, 1, 1, 1, 1,
0, 1, 0, 0, 0, 1, 0, 0, 0, 0,
1, 0, 1, 0, 1, 0, 0, 0, 0, 0,
1, 0, 1, 1, 1, 1, 1, 1, 1, 0,
1, 0, 0, 1, 0, 0, 1, 1, 1, 1,
1, 1, 1, 1, 1, 0, 0, 1, 0, 0,

1, 0, 1, 1, 1, 1, 1, 0, 0, 1,
1, 1, 1, 1, 0, 0, 1, 1, 1, 1,
0, 1, 1, 1, 0, 0, 1, 1, 0, 0,
1, 0, 0, 1, 0, 0, 1, 1, 1, 1,
0, 1, 1, 1, 1, 0, 1, 0, 1, 1,
0, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 0, 1, 0, 0, 1, 1, 1, 1, 1,
0, 1, 0, 1, 1, 0, 1, 0, 1, 0,
0, 0, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 0, 1, 0,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
0, 1, 1, 1, 1, 1, 1, 1, 1, 1,
0, 0, 0, 1, 1, 1, 1, 1, 0, 0,
0, 0, 1, 1, 1, 1, 1, 1, 1, 0,
0, 0, 0, 0, 0, 0, 0, 1, 0, 0,
0, 1, 1, 1, 0, 0, 1, 0, 1, 1,
1, 1, 1, 0, 1, 0, 0, 1, 1, 0,
0, 0, 1, 1, 0, 1, 0, 1, 0, 0,
0, 1, 0, 1, 0, 1, 0, 0, 0, 0,
0, 1, 1, 1, 1, 0, 0, 1, 0, 1,
0, 0, 1, 1, 1, 1, 1, 1, 1, 1,
0, 1, 0, 1, 1, 0, 1, 0, 0, 1,
1, 0, 1, 0, 1, 1, 0, 1, 1, 1,
0, 0, 0, 1, 0, 0, 1, 1, 1, 1,
0, 1, 1, 1, 1, 1, 1, 1, 1, 1,
0, 0, 0, 0, 0, 0, 1, 1, 1, 0,
0, 1, 0, 1, 0, 1, 1, 1, 1, 1,
1, 0, 0, 1, 1, 0, 1, 0, 0, 0,
0, 1, 0, 1, 0, 1, 1, 0, 1, 1,
0, 1, 0, 1, 1, 1, 1, 1, 0, 1,
0, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 0, 1, 1, 0, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
0, 0, 0, 1, 0, 1, 1, 1, 0, 1,
0, 0, 1, 1, 0, 0, 1, 0, 1, 1,
0, 0, 1, 1, 1, 0, 0, 0, 0, 1,
1, 0, 1, 1, 0, 0, 1, 1, 0, 1,
1, 0, 1, 1, 0, 0, 1, 0, 0, 1,
0, 0, 0, 1, 1, 1, 1, 0, 1, 0,
1, 1, 0, 0, 0, 1, 1, 1, 0, 0,
1, 1, 0, 1, 1, 0, 1, 1, 1, 0,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
0, 0, 0, 0, 1, 1, 1, 1, 1, 0,
0, 0, 0, 0, 0, 1, 0, 0, 0, 0,
0, 0, 1, 1, 0, 1, 1, 1, 1, 1,
1, 1, 1, 1, 0, 1, 1, 1, 1, 1,
1, 1, 0, 1, 0, 1, 1, 1, 0, 1,
0, 0, 0, 0, 0, 1, 0, 1, 1, 0,
1, 1, 1, 1, 0, 1, 1, 0, 1, 0,
1, 0, 1, 0, 0, 0, 0, 0, 1, 0,
1, 0, 0, 1, 0, 0, 1, 1, 1, 0,
0, 1, 1, 1, 0, 1, 1, 1, 1, 1,
0, 1, 1, 1, 0, 0, 1, 0, 1, 1,
0, 0, 1, 1, 1, 1, 1, 0, 0, 0,
0, 1, 1, 1, 0, 1, 1, 1, 1, 0,
0, 0, 1, 1, 1, 1, 0, 1, 0, 0,
1, 1, 1, 1, 0, 1, 0, 1, 0, 0,
0, 0, 1, 1, 0, 0, 1, 1, 0, 0,
1, 1, 1, 1, 1, 1, 1, 0, 1, 0,
0, 1, 0, 1, 0, 0, 1, 1, 0, 0,
0, 1, 1, 1, 1, 1, 1, 1, 0, 0,
0, 1, 0, 1, 0, 0, 1, 1, 1, 1,
0, 0, 0, 0, 0, 0, 0, 1, 0, 1,
0, 1, 0, 1, 0, 0, 1, 1, 1, 0,
0, 0, 1, 1, 0, 1, 1, 0, 1, 1,
1, 1, 1, 1, 0, 1, 1, 1, 1, 1,
0, 0, 1, 1, 0, 0, 1, 0, 1, 1,
0, 0, 1, 1, 0, 0, 1, 1, 1, 1,
0, 1, 1, 1, 0, 0, 1, 1, 0, 1,
1, 0, 1, 1, 0, 1, 1, 1, 1, 1,
0, 1, 1, 1, 0, 0, 1, 1, 0, 0,
0, 1, 1, 1, 1, 1, 1, 0, 0, 1,
0, 1, 0, 1, 1, 1, 1, 0, 1, 0,
0, 0, 1, 1, 0, 0, 1, 0, 1, 1,
1, 0, 1, 1, 0, 1, 1, 1, 1, 0,

```
1, 1, 1, 1, 0, 1, 1, 1, 1, 1,
0, 0, 0, 1, 1, 0, 0, 1, 1, 0,
0, 0, 0, 0, 1, 1, 0, 0, 1, 1,
1, 1, 0, 1, 0, 0, 1, 0, 0, 0,
0, 0, 1, 1, 0, 0, 1, 0, 1, 1,
0, 0, 0, 1, 0, 1, 0, 1, 1, 0,
0, 1, 0, 1, 1, 0, 1, 0, 0, 1,
0, 1, 1, 1, 0, 1, 1, 1, 0, 1,
1, 1, 1, 1, 0, 1, 0, 1, 1, 0,
0, 1, 1, 1, 0, 1, 1, 1, 1, 1,
1, 1, 1, 1, 0, 1, 1, 1, 1, 1,
0, 1, 1, 1, 1, 1, 0, 0, 1, 1,
0, 0, 0, 1, 1, 0, 1, 0, 0, 0,
0, 1, 1, 1, 1, 1, 1, 1, 0, 1,
1, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 1, 1, 1, 1, 1, 1, 1, 1,
0, 0, 0, 0, 0, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
0, 0, 1, 1, 0, 1, 1, 1, 0, 1,
1, 1, 1, 1, 0, 0, 0, 0, 1, 1,
0, 1, 1, 1, 0, 0, 1, 1, 0, 0,
0, 1, 0, 0, 0, 1, 0, 0, 0, 1,
0, 0, 1, 1, 0, 0, 1, 1, 1, 1,
0, 0, 0, 0, 0, 0, 1, 1, 1, 0,
0, 1, 0, 1, 1, 1, 1, 1, 1, 1,
0, 0, 1, 0, 0, 0, 0, 1, 1, 0,
1, 0, 1, 1, 0, 1, 1, 1, 1, 1),. Dim=c(250,10)))
```

الملحق (6)

ملف أوامر مستمد من برنامج WinBugs عندما يكون النموذج ثنائي المعلمة وعدد الفقرات 10 وحجم العينة 500 فرداً.

```
Model
}

# Read in individual item responses
for (i in 1:500)
}
for (j in 1:10)
}
x[i,j] <- response[i,j]
{
{

#Identify two-parameter logistic (2-PL ) model
for (i in 1:500)
}
for (j in 1:10)
}
p[i,j] <- exp(a[j]*(theta[i]-b[j]))/(1+exp(a[j]*(theta[i]-b[j])))
x[i,j] ~ dbern(p[i,j])
mx[i,j] <- 1-x[i,j]
{
#Specify prior for examinee parameters
theta[i] ~ dnorm(0,1)
{
#Specify priors for item parameters
for (j in 1:10)
}
a[j] ~ dlnorm(1,1. 649)
b[j] ~ dnorm(0,2)
{
#Generate replicate data for posterior predictive checks
for (i in 501:3000)
}
for (j in 1:10)
}
p[i,j] <- exp(a[j]*(theta[i]-b[j]))/(1+exp(a[j]*(theta[i]-b[j])))
r[i,j] ~ dbern(p[i,j])
mr[i,j] <- 1-r[i,j]
{
theta[i] ~ dnorm(0,1)
{
#compute odds ratios ofr relicate data
for (k in 2:10)
}
for (i in 1:(k-1))
}
or[k,i]<-
inprod(r[501:3000,i],r[501:3000,k])*inprod(mr[501:3000,i],mr[501:3000,k])/(inprod(mr[501:3000,i],r[501:3000,k])*inprod(r[5
01:3000,i],mr[501:3000,k]))
{
{

# End of model
```

Abstract

Shawawreh, Shadi Yousef, Accuracy of Estimating item Parameters Via Marginal Maximum Likelihood and Bayesian Methods in accordance with difference in Number of Items, Sample size and Logistic Models. Ph. D. Thesis, University of Yarmouk 2013 Supervised by: (Dr. Sari Sawaqed).

This study aimed to show which estimating method (maximum likelihood or Bayesian) is more accurate in estimating item parameters under any circumstances relating to the sample size, the number of items and the logistic model in use. To achieve the goal of the study the data has been generated with binary answers according to the logistic method in use (1-PL, 2-PL), size of sample (1500, 1000, 500, 250) and the number of items (40, 10).

After that, the item parameters were estimated under each circumstances of the previous (sample size, number of items and logistic model) using: maximum likelihood by (Bilog-MG 3) program and Bayesian by (WinBUGS v1.4) program.

The values of item parameters were analyzed according to both methods (maximum likelihood and Bayesian) using accuracy of estimating indicators such as accurate indicator Root Mean Square Error (RMSE), Relative Efficiency (RE) and correlation coefficient between the true Parameter value and the estimated one according to the methods of estimating (maximum likelihood and Bayesian).

Then, a t-test was made to compare between correlation coefficient of both estimating methods.

After the analyzing process is over, the results of the accurate indicator (RMSE) showed that Bayesian method is better in estimating and that was clear in the t-test results, which compared between correlation coefficients of the two methods of estimation. As the correlation coefficients of Bayesian

showed statistical significance on the level of significance (α=0.05) under most circumstances of the study.

Finally, based on the results of the study, we can say that the Bayesian method showed accuracy in estimating item parameters compared to maximum method especially in small samples or when the two logistic method (2-PL) is used.

Keywords: Maximum Likelihood, Bayesian method, estimating accuracy, sample size, number of items, Logistic model